

Université de Montréal

**Étude des anomalies du développement humain# un modèle d'analyse  
phénotypique**

Étude rétrospective d'une série d'autopsies fœtales

par

**Farideh Arbabzadeh**

Département de Pathologie, CHU Sainte-Justine

Faculté de médecine

Mémoire présentée à la Faculté des études supérieures, Udm

en vue de l'obtention du grade de M.Sc.

en Sciences biomédicales

option Recherche clinique

Juillet 2009

© Farideh Arbabzadeh, 2009

Université de Montréal  
Faculté des études supérieures

Cette thèse intitulée :

**Étude des anomalies du développement humain# un modèle d'analyse  
phénotypique**

Étude rétrospective d'une série d'autopsies fœtales

présentée par :  
Farideh Arbabzadeh

a été évaluée par un jury composé des personnes suivantes :

**Luc Oligny**, président- rapporteur  
**Jean-Christophe Fournet**, directeur de recherche  
**Chantal Bernard**, membre du jury

## Résumé

Depuis le début des années 90, le projet génome humain a permis l'émergence de nombreuses techniques globalisantes porteuses du suffixe –omique : génomique, transcriptomique, protéomique, épigénomique, etc....

L'étude globale de l'ensemble des phénotypes humains (« phénomène ») est à l'origine de nouvelles technologies constituant la « phénomique ». L'approche phénomique permet de déterminer des liens entre des combinaisons de traits phénomiques.

Nous voulons appliquer cette approche à l'étude des malformations humaines en particulier leurs combinaisons, ne formant des syndromes, des associations ou des séquences bien caractérisés que dans un petit nombre de cas.

Afin d'évaluer la faisabilité de cette approche, pour une étude pilote nous avons décidé d'établir une base de données pour la description phénotypique des anomalies fœtales. Nous avons effectué ces étapes :

- Réalisation d'une étude rétrospective d'une série d'autopsies de fœtus au CHU Sainte-Justine (Montréal, QC, Canada) entre 2001-2006
- Élaboration de trois thésaurus et d'une ontologie des anomalies développementales humaines
- Construction d'une base de données en langage MySQL

Cette base de données multicentrique accessible sur (<http://www.malformations.org>), nous permet de rechercher très facilement les données phénotypiques des 543 cas observés porteurs d'une anomalie donnée, de leur donner une description statistique et de générer les différents types d'hypothèses. Elle nous a également permis de sélectionner 153 cas de fœtus malformés qui font l'objet d'une étude de micropuce d'hybridation génomique comparative (aCGH) à la recherche d'une anomalie génomique.

**Mots-clés** : base de données MDB; anomalies développementales humaines; dysmorphologie; combinaisons malformatives; l'approche phénomique; ontologie phénotypique; anomalie génomique;

## Abstract

Since the early 90s, the Human Genome Project (HGP) has allowed the development of numerous worldwide techniques which carried the suffix “omic”: genomic, transcriptomic, proteomic, epigenomic, etc.... The global investigation of the sets of human phenotypes (phenome) is called phenomic. With phenomic studies we should be able to determine the links among similar phenotypic groups.

We wish to apply this approach to human dysmorphology, particularly malformation combinations, which form characteristic malformation associations, malformation sequences, malformation syndromes or malformation disorders only in a minority of cases.

As a graduate student research project, we decided to perform a retrospective study of the sets of pathology reports including 543 fetuses autopsied in the Department of Pathology of CHU Sainte-Justine (Montreal, QC, Canada) between 2001 and 2006.

We have established an open Malformation Database (MDB) which can be accessed at <http://www.malformations.org>. To achieve this, we conducted the following steps:

- Realization of a retrospective study of fetopathology reports for fetal malformations.
- Development of an ontology along with three thesauruses of human developmental anomalies.
- Implementation of these thesauruses and ontology in the MySQL system.

This hypothesis-generating database allows us to easily retrieve the fetal cases (phenotypic data) with anomalies, calculate the frequencies of these anomalies, and evaluate the feasibility of the phenomic approach to human dysmorphogenesis. We were able as well to select 153 cases of malformed fetuses which will be the subject of aCGH array study for genomic research of human anomalies.

**Keywords:** malformation database; MDB; human developmental anomalies; dysmorphology; malformation combinations; phenomic approach; phenotype ontology.

## Table des matières

Résumé.....	iii
Abstract .....	iv
Table des matières.....	v
Liste des tableaux.....	viii
Liste des figures .....	ix
Liste des sigles et des abréviations .....	x
Remerciements.....	xii
Introduction.....	1
Chapitre I: Contexte de la recherche.....	3
I.A. Anomalies développementales, malformations, syndromes malformatifs.....	4
I.A.1. Malformations primaires.....	4
I.A.2. Malformations secondaires .....	5
I.A.3. Défauts d'organogenèse .....	5
I.A.4. Malformations isolées ou multiple - combinaisons malformatives .....	6
I.A.4.a. Séquence malformative .....	6
I.A.4.b. Association malformative .....	7
I.A.4.c. Syndrome malformatif.....	7
I.A.5. Anomalie de champ développemental.....	8
I.A.6. Nosologie des syndromes malformatifs.....	8
I.B. Étiologie des anomalies développementales .....	9
I.B.1. Causes génétiques .....	9
I.B.2. Anomalies chromosomiques .....	9
I.B.3 Causes toxiques.....	12
I.B.4. Causes infectieuses .....	13
I.B.5. Causes maternelles.....	13
I.B.6. Causes multifactorielles .....	14
I.C. Le stockage et le traitement des données en fœtopathologie .....	14
I.C.1. POSSUM.....	15
I.C.2. La Winter-Baraitser Dysmorphology Database (WBDD) .....	15
I.C.3. CDDM Computerized Dysmorphology Database Mackay) .....	16
I.C.4. OMIM (On-line Mendelian Inheritance in Man).....	16
I.C.5. Orphanet.....	17
I.C.6. humpath.com.....	17
I.C.7. eMedicine.....	17
I.D. La nécessité d'une ontologie des anomalies développementales humaines : terminologie, thésaurus, ontologie .....	19
I.D.1. Thésaurus .....	19
I.D.2. Ontologies.....	22
I.D.3. Ontologies phénotypiques.....	23
I.E. Phénotype, phénomène et phénomique.....	24
I.E.1. Bases de données phénotypiques ou phénomiques .....	25
Chapitre II. Objectifs de notre travail .....	29
Chapitre III. Matériel et méthodes .....	31
III.A. Généralités sur l'examen fœtopathologique .....	32
III.B. L'autopsie fœtale au CHU Sainte- Justine.....	33

III.B.1. Photographies .....	33
III.B.2. Radiographies .....	33
III.B.3. Culture de fibroblastes.....	34
III.B.4. Examen morphologique.....	34
III.B.5. Examen histologique .....	34
III.B.6. Examens complémentaires .....	35
III.B.7. Examens de microbiologie et de virologie .....	35
III.B.8. Microscopie électronique .....	35
III.B.9. Softpath .....	35
III.C. Procédures de fonctionnement de la base de données MDB .....	36
III.C.1. Access.....	36
III.C.2. Une base de données en langage MYSQL .....	37
III.C.3. Le type de données recueillies.....	38
III.C.4. Le niveau d'identification des données et échantillons (codées, anonymisées) .....	38
III.C.5. Le lieu de conservation des données .....	38
III.C.6. Convention régissant l'utilisation des données du matériel biologique .....	39
III.C.7. L'approbation du comité d'éthique de la recherche (CÉR).....	39
Chapitre IV. Les étapes de notre travail.....	41
IV.A. Élaboration d'une base de données des malformations humaines.....	42
IV.B. Élaboration d'une ontologie des malformations fœtales .....	42
IV.B.1. Construction d'un thésaurus des anomalies développementales humaines ....	43
IV.B.2. Construction d'un thésaurus des combinaisons malformatives humaines .....	45
IV.B.3. Construction d'un thésaurus des maladies malformatives humaines.....	47
IV.B.4. Bilan .....	48
IV.C. Saisie des données phénotypiques des autopsies de fœtus malformés .....	48
IV.C.1. Élaboration d'une fiche de saisie des données phénotypiques.....	49
IV.C.2. Étude rétrospective de la série d'autopsies des fœtus du CHU Sainte-Justine entre 2001 et 2006#Saisie des données fœtopathologiques dans la base de données MDB.....	49
IV.D. Publication de la base de données sur le réseau Internet .....	51
Chapitre V : Discussion et perspectives.....	52
V.A. Automatisation de la collecte de données.....	53
V.B. Les usages des bases de données phénotypiques et phénomiques.....	53
V.B.1. Importance des statistiques et des données quantitatives .....	54
V.B.2. Les corrélations génotype- phénotype.....	54
V.B.3. Distance phénotypique .....	54
V.B.4. Agrégation phénotypique et familles de maladies.....	56
V.B.5. Proximité phénotypique - Proximité protéique .....	58
V.B.6. Agrégation phénotypique .....	58
V.B.7. Réseau phénotypique.....	58
V.B.8 Réseau phénotypique et réseau protéique.....	59
V.C. Phéno et nosologie .....	60
V.C.1. Une ontologie des phénotypes humains .....	61
V.C.2. Aide au diagnostic .....	62
V.C.3. Description de nouvelles combinaisons malformatives .....	63
V.C.4. Visualisation des données.....	63
V.C.5. Corrélations génotype phénotype .....	64

Conclusion .....	65
Bibliographie.....	67
Annexes.....	xiii

## Liste des tableaux

<b>Tableau 1 :</b> Quelques exemples de maladies malformatives monogéniques ou oligogéniques .....	11
<b>Tableau 2 :</b> Les caractéristiques des bases de données phénotypiques chez l'homme ...	18



## Liste des figures

<b>Figure 1 :</b>	Schéma illustrant les différents types d'erreurs de la morphogenèse.....	5
<b>Figure 2 :</b>	Dendrogramme par l'entrée « syndrome de Bardet-Biedl ».....	55
<b>Figure 3 :</b>	La famille des « ciliopathies ».. .....	57
<b>Figure 4 :</b>	Interactions entre un réseau de maladies et un réseau de gènes ou de protéines : le réseau phénomène-génome. ....	60
<b>Figure 5 :</b>	Une ontologie des phénotypes humains.....	61

## Liste des sigles et des abréviations

- **AD**, Anomalies Développementales
- **ADICAP**, Association pour le Développement de l'Informatique en Cytologie et en Anatomie Pathologique). <http://www.adicap.asso.fr/>
- **CDDM**, <http://www.ncbi.nlm.nih.gov/pubmed/8178639>
- **CÉR**, Comité d'Éthique de la Recherche
- **CGH**, array, micropuce d'hybridation génomique comparative
- **CSIT**, Centre de Support Informatique
- **DFP**, Developmental Field Defect
- **DSP**, Directeur des Services Professionnels  
[http://dsp.aqesss.qc.ca/997/Site\\_DSP.aqesss](http://dsp.aqesss.qc.ca/997/Site_DSP.aqesss)
- **EMBL**, European Molecular Biology Laboratory. <http://www.embl.de/>
- **Ensembl**, <http://uswest.ensembl.org/index.html>
- **GO**, Gene Ontology. <http://www.geneontology.org/>
- **HGMD**, Human Gene Mutation Database <http://www.hgmd.cf.ac.uk/ac/index.php>
- **HGP**, the Human Genome Project
- **HL7**, Health Level Seven International. <http://www.hl7.org/>
- **HPO**, Human Phenotype Ontology. <http://www.human-phenotype-ontology.org/>
- **HSJ**, CHU Sainte Justine
- **IHE**, Integrating the Healthcare Enterprise®. <http://www.ihe.net/>
- **IHTSDO**, International Health Terminology Standards Development Organisation.  
<http://www.ihtsdo.org/>
- **MausDB**, the mouse management system of the German Mouse Clinic  
<http://146.107.3.20/cgi-bin/mausdb/mausdb.cgi>
- **MDB**, Malformation DataBase. <http://www.malformations.org>
- **MEDLINE**, Medical Literature Analysis and Retrieval System Online  
[http://www.nlm.nih.gov/databases/databases\\_medline.html](http://www.nlm.nih.gov/databases/databases_medline.html)
- **MeSH**, Medical Subject Headings. <http://www.nlm.nih.gov/mesh/>
- **MPD**, Mouse Phenome Database. <http://phenome.jax.org/>
- **MPO**, Mammalian Phenotype Ontology. <http://www.informatics.jax.org/>

- **MySQL**, My Structured Query Language. <http://www.mysql.com/>
- **NLM**, National Library of Medicine. <http://www.nlm.nih.gov/>
- **OBO**, Open Biomedical Ontologies. <http://obofoundry.org/>
- **OLS**, Ontology Lookup Services. <http://www.ebi.ac.uk/ontology-lookup/>
- **OMIM**, Online Mendelian Inheritance in Man  
<http://www.ncbi.nlm.nih.gov/sites/entrez?db=omim>
- **PathLex**, lexique unique d'anatomopathologie lancé par l'IHE et HL7  
<http://www.hl7.org/Library/Committees/anatomicpath/>
- **PATO**, The Phenotype Quality Ontology, anciennement nommé « Phenotype Attribute and Trait Ontology »  
[http://www.bioontology.org/wiki/index.php/PATO:Phenotypes\\_in\\_OWL](http://www.bioontology.org/wiki/index.php/PATO:Phenotypes_in_OWL)
- **PMSI**, Patients' Medical Supporting Informations
- **POSSUM**, Pictures Of Standard Syndromes and Undiagnosed Malformations  
<http://www.possum.net.au/>
- **RGD**, Rat Genome Database. <http://rgd.mcw.edu/>
- **SGBD**, Système de Gestion de Base de Données
- **SNOMED**, Systematized Nomenclature of Medicine  
[http://www.nlm.nih.gov/research/umls/Snomed/snomed\\_main.html](http://www.nlm.nih.gov/research/umls/Snomed/snomed_main.html)
- **SQL**, Structured Query Language. <http://www.sql.su/>
- **WHO-ICD**, classification internationale des maladies publiée par l'Organisation Mondiale de la Santé. <http://www.dimdi.de/static/en/index.html>
- **WBDD**, the Winter-Baraitser Dysmorphology Database  
<http://www.lmdatabases.com/>

## Remerciements

Mes sincères remerciements vont au Docteur Luc Oligny-le chef de département de pathologie du CHU Sainte-Justine qui a contribué de près ou de loin à la réalisation de ce travail en particulier. Je tiens à remercier aussi :

- Le Dr Jean-Christophe Fournet, directeur de recherche pour ses idées novatrices.
- Le Département de Pathologie du CHU Sainte Justine
- M. Philippe Sawicki pour avoir si gentiment produit le serveur de base de données relationnelle MySQL

## **Introduction**

Depuis 20 ans, les progrès des techniques d'imagerie anténatale ont révolutionné le dépistage précoce des malformations. La pratique croissante des interruptions thérapeutiques de grossesse (ITG) est à l'origine de la description de dizaines de milliers de phénotypes malformatifs peu compris.

## **Chapitre I : Contexte de la recherche**

## **I.A. Anomalies développementales, malformations, syndromes malformatifs**

Les anomalies congénitales sont responsables d'une partie importante de la morbidité et de la mortalité infantiles ainsi que de la mortalité fœtale, d'où l'importance de recueillir des données épidémiologiques élémentaires sur ces anomalies. Environ 350 000 enfants naissent chaque année au Canada; 2 % à 3 % de ces enfants sont atteints d'une anomalie congénitale<sup>1</sup>. Le plus souvent, les parents de ces enfants n'ont aucun antécédent familial d'anomalies congénitales, et ne sont porteurs d'aucun facteur de risque connu.

Les anomalies du développement humain, ou anomalies développementales (AD) sont des anomalies de la conformation d'un tissu ou d'un organe ou d'une partie plus étendue de l'organisme lors de son ontogenèse ou de son développement. La gravité des malformations est extrêmement variable. Elle va de la malformation mineure, simple variante du normal, sans pathologie fonctionnelle jusqu'aux malformations sévères incompatibles avec la vie.

Une nomenclature internationale, basée principalement sur le moment où survient l'anomalie, établit les diverses catégories des anomalies développementales rencontrées chez le fœtus (Spranger, Benirschke et al. 1982). Elles se divisent en:

### **I.A.1. Malformations primaires**

Les **malformations primaires**, ou **dysgénésie**, résultent d'un événement génétiquement déterminé. Elles sont aussi appelées « **malformations intrinsèques** ».

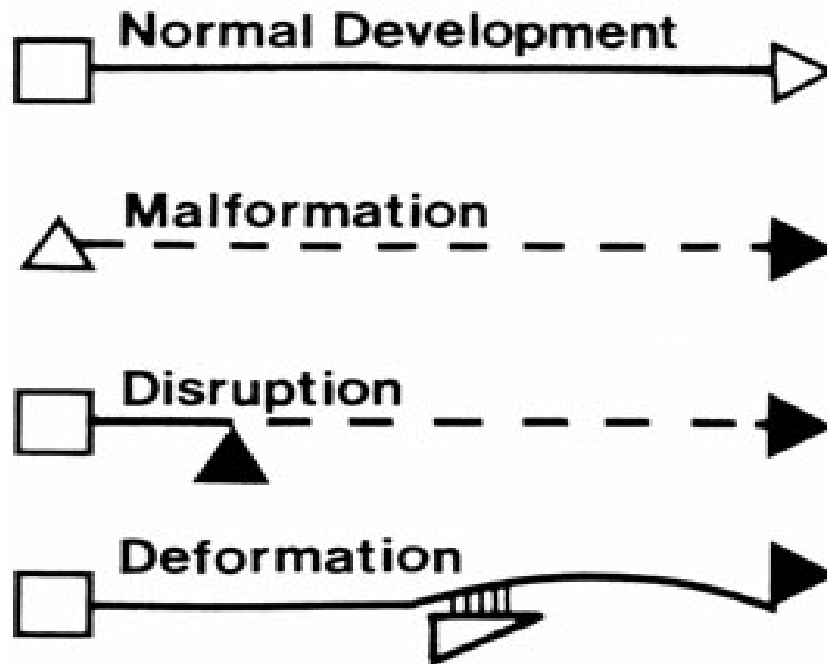
---

<sup>1</sup> <http://www.phac-aspc.gc.ca/publicat/cac-acc02/index-fra.php>



### I.A.2. Malformations secondaires

Les **malformations secondaires** résultent d'un **facteur extrinsèque** perturbant les processus normaux du développement; soit une perturbation de la formation normale d'une structure (**déformation**), soit une lésion secondaire d'un organe ou d'une structure déjà formés (**disruption**).



**Figure 1 :** Schéma illustrant les différents types d'erreurs de la morphogenèse par Judith G. Hall. D'après figure 5-12, page 132 de *Developmental Pathology of the Embryo and Fetus*. James E. Dimmick, et Dagmar K. Kalousek (Éditeurs), 875 pp, Philadelphie, J.B. Lippincott, 1992.

### I.A.3. Défauts d'organogenèse

Les **défauts d'organogenèse** sont un dernier type de malformation. Elles sont causées par l'anomalie de développement d'un organe (Spranger, Benirschke et al. 1982). Elles sont souvent désignées par le terme de « dysplasie » ou de « dysgénésie », signifiant « anomalie

de développement » : dysplasie rénale, dysgénésie gonadique. D'autres peuvent intéresser la migration d'un type cellulaire dans un organe donné, par exemple les anomalies de la migration neuronale dans le cerveau.

#### **I.A.4. Malformations isolées ou multiple - combinaisons malformatives**

Une malformation peut être isolée ou multiple. Une **malformation isolée** peut par exemple être une polydactylie isolée ou une fente labiale isolée. Inversement, les anomalies développementales élémentaires peuvent aussi être multiples et se grouper de façon non-aléatoire pour former des **combinaisons malformatives**.

Ces combinaisons peuvent être **monotopiques** (avoir une seule localisation) et **polytopiques** (avoir plusieurs localisations). Certaines de ces combinaisons présentent des régularités et des spécificités permettant de les reconnaître. Les combinaisons malformatives sont identifiées par le nom de leurs descripteurs (éponyme; comme le syndrome de Meckel) ou par un acronyme formé à partir de la première lettre de chaque malformation associée (comme les associations VACTERL ou CHARGE).

Les combinaisons malformatives peuvent être des **séquences malformatives**, des **associations malformatives**, ou des **syndromes malformatifs**.

##### ***I.A.4.a. Séquence malformative***

La **séquence malformative** est une combinaison malformative formée par une cascade d'événements entraînés par une cause initiale (par exemple, la séquence d'oligohydromnios ou « de Potter », la séquence de Prune-Belly de la distension vésicale).

#### ***I.A.4.b. Association malformative***

Une **association malformative** est une association de malformations sans lien fonctionnel ou topographique entre elles. Elles sont le plus souvent désignées par des acronymes.

Il s'agit par exemple des associations :

- CHARGE (Coloboma, Heart malformations, choanal Atresia, Retardation of growth, Genital and Ear anomalies)
- MURCS (Müllerian duct aplasia, Unilateral Renal agenesis, Cervicothoracic Somite anomalies)
- VATER ou VACTERL (Vertebral anomalies, Anal atresia, Cardiac malformations, TracheoEsophageal fistula, Renal anomalies, Radial dysplasia and other limb defects)

#### ***I.A.4.c. Syndrome malformatif***

Un **syndrome malformatif** est une combinaison reconnaissable d'anomalies développementales (en général des malformations), qui sont associées par une même cause (étiologie) : une anomalie chromosomique, une mutation d'un gène, une infection, une exposition à un toxique tératogène. La séparation entre « association » et « syndrome » est ainsi de moins en moins pertinente car il est maintenant connu que les associations sont souvent de cause monogénique, par exemple les mutations de *CHD7* dans l'association CHARGE.

### **I.A.5. Anomalie de champ développemental**

Une combinaison malformative peut être due à un phénomène dysmorphogénétique anormal, dans un champ morphogénétique donné. Il s'agit alors d'une « anomalie de champ développemental » (Developmental Field Defect, DFD) (Martinez-Frias, Frias et al. 1998). Ainsi, la plupart des anomalies développementales survenant au cours de la blastogenèse sont polytopiques, et impliquent deux ou plusieurs champs progénitures (descendances), par exemple, le champ acrorénal, gastromélique, ou splénomélique. Elles sont à l'origine de syndromes malformatifs très sévères touchant plusieurs organes.

### **I.A.6. Nosologie des syndromes malformatifs**

Depuis le début du vingtième siècle, la régularité et la répétition de ces combinaisons non-aléatoires a permis d'identifier des centaines de syndromes malformatifs, indexés dans OMIM parmi les autres maladies génétiques mendéliennes. Le chevauchement des phénotypes a déclenché d'importants débats pour la création de nouvelles entités nosologiques ou la fusion d'entités voisines. Ce débat a été appelé « The lumpers and splitters debate » (McKusick 1969).

Si des centaines de syndromes malformatifs ont été décrits, donnant naissance à une science, appelée la « syndromologie » (Fitch 1977), en pratique diagnostique quotidienne, la plupart des combinaisons observées ne correspondent pas à des entités bien définies.

## **I.B. Étiologie des anomalies développementales**

De nombreux facteurs ont été identifiés à l'origine ou comme pouvant favoriser la constitution d'une anomalie congénitale en général, mais il est souvent difficile de retrouver avec précision le facteur en cause dans une malformation congénitale en particulier.

### **I.B.1. Causes génétiques**

Les **causes génétiques** peuvent être liées soit à une anomalie chromosomique (maladie malformative chromosomique), soit à une anomalie d'un ou plusieurs gènes (maladie malformative monogénique-oligogénique et polygénique). Les causes génétiques rendent compte d'environ 25 % des anomalies congénitales (Nelson et Holmes 1989).

### **I.B.2. Anomalies chromosomiques**

Les **anomalies chromosomiques** sont des causes très fréquentes de malformations. Les plus fréquentes sont des trisomies de chromosomes acrocentriques (trisomie 13, trisomie 18, trisomie 21), qui sont à l'origine d'importants syndromes malformatifs. D'autres syndromes malformatifs sont associés à des microdélétions chromosomiques, ou à d'autres réarrangements structuraux, dont l'expressivité est variable (Veltman et Brunner 2010). Le développement des techniques de micropuce d'hybridation génomique comparative (CGH array) est à l'origine de la description de nombreuses microdélétions causant des malformations (Shaffer, Bejjani et al. 2007; Shaffer, Theisen et al. 2007; Slavotinek 2008). Ces techniques retrouvent environ 10% d'anomalies chromosomiques dans les enfants porteurs de malformations. (Stankiewicz et Beaudet 2007; Hillman, Pretlove et al. 2010).

Les anomalies d'un gène donné peuvent être à l'origine d'une **maladie malformative monogénique**, également dite **maladie mendélienne**. Comme pour les autres maladies mendéliennes, la transmission des syndromes malformatifs familiaux peut être récessive, dominante ou lié à l'X. Il peut s'agir de gènes codant pour des facteurs de transcription, des molécules d'adhésion et plus rarement des enzymes catalysant ces réactions.

**Tableau 1 :** Quelques exemples de maladies malformatives monogéniques ou oligogéniques

- Kabuki mask syndrome (MIM.147920): *MLL2*
- Joubert syndrome (MIM.213300)
  - JBTS1(MIM.213300) : mutations de *INPP5E*
  - JBTS2 (MIM.608091) : mutations de *TMEM216* en 11q13
  - JBTS3 (MIM.608629) : mutations de *AHI1* en 6q23
  - JBTS4 (MIM.609583) : mutations de *NPHP1* (607100) en 2q13
  - JBTS5 (610188) : mutations de *CEP290*, aussi appelé *NPHP6*(610142) en 12q21.32
  - JBTS6 (610688) : mutations de *TMEM67* (609884) en 8q21
  - JBTS7 (611560) : mutations de *RPGRIP1L* (610937) en 16q12.2
  - JBTS8 (612291) : mutations de *ARL13B* (608922) en 3q11.2
  - JBTS9 (612285) : mutations de *CC2D2A* (612013) en 4p15.3
  - JBTS10 (300804) : mutations de *CXORF5* (300170) en Xp22.3
- Syndrome de Meckel
  - MKS1 (249000) : mutations de *MKS1* en 17q23
  - MKS2 (603194) : mutations de *TMEM216* en 11q13
  - MKS3 (607361) : mutations de *TMEM67* en 8q
  - MKS4 (611134) ; mutations de *CEP290* (610142) en 12q
  - MKS5 (611561) ; mutations de *RPGRIP1L* (610937) en 16q12.2
  - MKS6 (612284) : mutations de *CC2D2A* (612013) en 4p15
- Fente labiale, avec ou sans fente palatine: mutations *MAFB* ou *ABCA4*
- Syndrome de Schinzel-Giedion : mutations de *SETBP1*
- Anémie de Fanconi (MIM;227650): mutations des gènes *FANCA*, *FANCB*, *FANCC*, *FANCD1*, *FANCD2*, *FANCE*, *FANCF*, *FANCG*, *FANCI*, *FANCJ*, *FANCL*, *FANCM*, *FANCN*, et *RAD51C*
- Syndrome de Noonan (MIM.163950): mutations de *PTPN11* en 12q24.1
- Dysplasie généralisée des vaisseaux lymphatiques (MIM.235510): mutations de *CCBE1*
- Brachydactylie-anonychie (MIM.106995): mutations de *SOX9*
- Syndrome de Smith-Lemli-Opitz (MIM.270400): mutations de *DHCR7*

### **I.B.3 Causes toxiques**

Les **causes toxiques** sont des agents chimiques qui exercent un effet néfaste sur l'embryon ou le fœtus entre la fécondation et la naissance. Le rapport entre les risques tératogènes résultant de la plupart des expositions environnementales maternelles et les anomalies congénitales n'est pas encore bien défini (Shaw, Wasserman et al. 1999). Les effets des expositions paternelles le sont encore moins (Lin, Hwang et al. 1998; Trasler and Doerksen 1999).

L'agent tératogène le plus couramment consommé est l'alcool. Le syndrome d'alcoolisme fœtal (SAF) est, au Canada, l'une des principales causes d'anomalies congénitales et de retards de développement évitables chez l'enfant (Luke 1977; Krous 1981; Chudley, Kilgour et al. 2007). Quelques études épidémiologiques ont démontré que les incinérateurs de déchets provoquent la naissance d'enfants malformés (Boyle, Johnson et al. 2004). Les auteurs de ces études estiment que globalement, des risques significatifs pour les populations exposées sont observés pour deux types de malformations : les anomalies chromosomiques (Vrijheid, Dolk et al. 2002) et les autres malformations majeures.

Plusieurs médicaments ont également fait la preuve de leur pouvoir tératogène. Les principaux médicaments tératogènes sont la thalidomide, les antagonistes de l'acide folique, les anticonvulsivants (Dilantin, Tégrétol), les dérivés de la coumarine et les rétinoïdes (Accutane).

L'environnement joue probablement un rôle critique dans la genèse des malformations (Ferencz, Loffredo et al. 2000; Dolk 2004). Des recherches récentes ont signalé un risque accru de malformations anatomiques et d'anomalies chromosomiques lié à la pollution



atmosphérique (Monteleone-Neto, Brunoni et al. 1985; Gold 1990; 2002; 2002; Ritz, Yu et al. 2002; Hansen, Barnett et al. 2009) et à la proximité des sites de déchets dangereux (Boyle, Johnson et al. 2004; Kloppenborg, Brandt et al. 2005; Palmer, Dunstan et al. 2005; Elliott, Richardson et al. 2009). Les résultats des recherches menées sur l'exposition maternelle aux pesticides (Shaw, Wasserman et al. 1999; Kumar 2002), et sur les produits dérivés du trihalométhane ont également démontré leur potentiel tératogène (Klotz and Pyrch 1999; Dodds and King 2001).

#### **I.B.4. Causes infectieuses**

Les **causes infectieuses**: Parmi les agents infectieux qui peuvent être transmis au fœtus et avoir sur lui un effet néfaste figurent entre autres le parvovirus B-19, le virus de la rubéole, le cytomégalovirus, le virus de la varicelle, le virus de l'herpès et les agents de la toxoplasmose et de la syphilis.

#### **I.B.5. Causes maternelles**

Les **causes maternelles**: L'âge de la mère constitue aussi un facteur de risque d'anomalies congénitales et plus particulièrement d'anomalies chromosomiques. Les affections maternelles qui peuvent contribuer à l'augmentation des risques d'anomalies congénitales sont les suivantes : obésité en relation avec le diabète sucré (Loeken 2006) et les troubles thyroïdiens maternels (hyperthyroïdie et hypothyroïdie) (Khoury, Becerra et al. 1989; Browne, Rasmussen et al. 2009).

### **I.B.6. Causes multifactorielles**

Les **causes multifactorielles** : Certaines des anomalies congénitales courantes sont réputées d'origine multifactorielle, et leur apparition tient de l'influence combinée de facteurs environnementaux et génétiques. Une anomalie développementale est réputée d'origine multifactorielle (ou polygénique) lorsqu'elle résulte de l'influence combinée d'un certain nombre de gènes et de facteurs environnementaux qui nuisent au développement embryologique normal. L'hérédité multifactorielle n'est envisagée que lorsqu'une transmission mendélienne est exclue et qu'il n'y a aucune évidence d'anomalie génétique à transmission mendélienne ou non mendélienne, d'aberration moléculaire ou chromosomique. Un exemple de cause multifactorielle est donné par les anomalies de fermeture du tube neural, pour lesquelles sont suspectées une carence d'apport maternel en folate dans un contexte génétique particulier (Neumann, Frankel et al. 1994).

### **I.C. Le stockage et le traitement des données en fœtopathologie**

Environ 4000 maladies génétiques ont été décrites, dont plusieurs centaines de maladies malformatives. Elles associent des milliers de caractères phénotypiques encore mal recensés et caractérisés. La quantité d'information à traiter est donc très importante et rend nécessaire la mise en œuvre de bases de données. Or, peu de bases de données sont encore disponibles en fœtopathologie, et elles sont souvent privées et non accessible sans avoir payé.

### **I.C.1. POSSUM**

La base de données iconographique **POSSUM**<sup>2</sup> (Pictures Of Standard Syndromes and Undiagnosed Malformations) (Bankier and Keith 1989; Pelz, Arendt et al. 1996) a été développée au Murdoch Childrens Research Institute, du Royal Children's Hospital de Victoria en Australie. Elle contient plus de 3000 syndromes incluant des anomalies chromosomiques et métaboliques. **POSSUM** est une base de données fermée et privée, commercialisée sous forme de CD puis d'un site web à accès restreint. POSSUM contient 3167 syndromes et 1115 traits phénotypiques. 924 syndromes sont associés à un gène muté et 986 gènes de maladies (disease genes) y sont recensés.

### **I.C.2. La Winter-Baraitser Dysmorphology Database (WBDD)**

Appelée également London Dysmorphology Database, la Winter-Baraitser Dysmorphology Database (WBDD)<sup>3</sup> contient les informations sur plus de 4450 anomalies du développement humain, incluant les retards mentaux ainsi que 44000 références bibliographiques (Brandl and Grimm 1987; Winter and Baraitser 1987; Baraitser and Winter 1991; Guest, Evans et al. 1999; Fryns and de Ravel 2002).

Cette base de données privée est également commercialisée sur le web (environ \$770).

---

<sup>2</sup> <http://www.possum.net.au/>

<sup>3</sup> <http://www.lmdatabases.com/>

### I.C.3. CDDM (Computerized Dysmorphology Database Mackay)

La base de données informatisée de Mackay en dysmorphologie (CCMD)<sup>4</sup> est une base de données dysmorphologique, gérée par le programme Clipper (Hsu, Lin et al. 1994). En 1994, 278 maladies malformatives avec leurs caractéristiques cliniques et les références associées étaient stockées dans cette base de données. Il n'y a pas eu de rehaussement depuis 1994.

### I.C.4. OMIM (On-line Mendelian Inheritance in Man)

OMIM<sup>5</sup> est une base de données dédiée aux maladies génétiques mendéliennes (maladies monogéniques), dont des maladies génétiques malformatives (Schorderet 1991; Amberger, Bocchini et al. 2009). Développée par Victor McKusick à partir de son livre **Mendelian Inheritance in Man**, OMIM est un catalogue des gènes humains et des maladies génétiques mendéliennes, dont certaines sont à l'origine de malformations. Hébergée par le NCBI, la base de données OMIM permet d'accéder gratuitement à des informations cliniques, génétiques et moléculaires et de trouver des informations médicales précises sur de très nombreuses maladies génétiques. La base OMIM contient 4779 syndromes selon HPO (Human Phenotype Ontology) et 5948 selon MimMiner. Elle contient 8275 caractères ou traits phénotypiques selon HPO. Les synopsis cliniques des entrées OMIM ont été récemment convertis en une forme structurée par le projet HPO conduit par Peter Robinson à l'Hôpital de la Charité de Berlin.

---

<sup>4</sup> <http://www.ncbi.nlm.nih.gov/pubmed/8178639>

<sup>5</sup> <http://www.ncbi.nlm.nih.gov/omim>

### **I.C.5. Orphanet**

Orphanet<sup>6</sup> n'est pas une base de données proprement dite mais un serveur d'information pour tous publics sur les maladies rares et orphelines (Ayme 2000; Ayme et Schmidtke 2007). C'est une encyclopédie en ligne et en plusieurs langues de ces pathologies, associée à un répertoire des ressources disponibles au grand public et aux scientifiques. Une partie de ces maladies rares et orphelines sont des maladies malformatives. Orphanet a été développé par le Pr Ségolène Aymé (Université Pierre et Marie Curie, Paris 6). Orphanet contient des informations sur 7435 syndromes. Il utilise une définition clinique des syndromes, ce qui est différent d'OMIM, qui base ses fiches sur des loci (locus morbide). En 2009, Orphanet recensait 569 gènes de maladies (disease genes).

### **I.C.6. humpath.com**

Le site web humpath.com<sup>7</sup> est dédié à la pathologie pour l'étude morphologique et moléculaire des maladies humaines, créé par le Dr Jean-Christophe Fournet, pathologiste pédiatrique au CHU Sainte-Justine de Montréal. Cette étude se place à tous les niveaux d'organisation du corps humain (molécules, cellules, tissus, organes, systèmes, corps). Une grande partie du site est consacré aux anomalies développementales et aux maladies malformatives, et associe de l'iconographie.

### **I.C.7. eMedicine**

eMedicine<sup>8</sup> est une base de données médicale en ligne créée en 1996 par Scott Plantz et

---

<sup>6</sup> <http://www.orpha.net/consor/cgi-bin/index.php>

<sup>7</sup> <http://www.humpath.com/>

<sup>8</sup> <http://emedicine.medscape.com/>

Richard Lavelly,. Elle fut rachetée par WebMD en janvier 2006 et s'adresse aux médecins et autres professionnels de la santé. Environ 10000 médecins et auteurs y contribuent, couvrant 7000 maladies et syndromes. Cette base, qui est en constante mise à jour, propose des liens vers les différentes spécialités médicales. Comme pour humpath.com, une partie du site est consacrée aux anomalies développementales et aux maladies malformatives.

**Tableau 2 :** Les caractéristiques des bases de données phénotypiques chez l'homme d'après Oti, M., M. A. Huynen, et al. (2009). "The biological coherence of human phenome databases." Am J Hum Genet 85(6): 801-808

	<b>Orphanet</b>	<b>POSSUM</b>	<b>OMIM-HPO</b>	<b>OMIM-MimMiner</b>
Nombre des Syndromes	2070 (7435)	3167	4779	5948
Nombre des Caractères d'ontologie	864	1115	8275	1368
Profondeur des Caractères d'ontologie : max. (médian)	4 (2)	2 (2)	13 (6)	15 (5)
Nombre médian de caractères/ par syndromes	13(25)	22(34)	7(20)	8(22)
Nombre de syndromes reliés à un gène	668	924	2053	2055
Nombre de gènes de maladies	1038	986	2019	1937

Le nombre entre parenthèses indique le nombre médian de caractères, (voir page xiv- annexe 1).  
 OMIM-HPO, (voir page 61- V.C.1. Une ontologie des phénotypes humains).  
 OMIM- MimMiner, (voir page 54- V.B.3. Distance phénotypique).

Ces bases de données entretenues manuellement posent des problèmes méthodologiques. Ainsi, la base de données OMIM gère la plus vaste collection des maladies humaines, mais sa description phénotypique (l'annotation des données phénotypiques) est mal structurée et ne convient pas à une analyse informatique automatisée (l'analyse phénotypique). La construction des vraies bases de données phénotypiques est dépendante de la disponibilité de thésaurus et d'ontologies dédiées aux phénomènes humains ou animaux. Des bases de données informatisées, comme PhenoGO<sup>9</sup>, sont dans ce domaine plus appropriées (Lussier, Borlawsky et al. 2006; Lussier YA, Liu Y. 2007; Sam, Mendonca et al. 2009). (Voir page xv- annexe 2 et page 25- I.E.1. Bases de données phénotypiques ou phénomique).

## **I.D. La nécessité d'une ontologie des anomalies développementales humaines : terminologie, thésaurus, ontologie**

Le développement de techniques informatiques pour la gestion et la connaissance des malformations humaines nécessite au préalable la constitution de thésaurus et d'ontologies spécialisés et précis.

### **I.D.1. Thésaurus**

Un thésaurus est un glossaire hiérarchisé utilisant un vocabulaire normalisé sur la base de termes génériques et de termes spécifiques à un domaine. Il est constitué d'un ensemble structuré de termes pouvant être utilisé pour l'indexation de documents dans une banque de données bibliographiques ou dans un catalogue de centre de documentation. Depuis une vingtaine d'années, plusieurs thésaurus médicaux ont été mis au point, dans lesquels sont

---

<sup>9</sup><http://www.phenogo.org/>

inclus une ontologie partielle des malformations : SNOMED, IHTSDO, ADICAP, WHO - ICD, POSSUM.

### ***1.D.1.a. SNOMED***

SNOMED<sup>10</sup> (Systematized Nomenclature of Medicine) est le système de codage des maladies humaines, des syndromes lésionnels et des lésions développé par l'International Health Terminology Standards Development Organisation (IHTSDO). À but lucratif et onéreux, protégé par des droits d'auteurs, il contient peu d'entrées en pathologie fœtale et malformative.

### ***1.D.1.b. IHTSDO***

IHTSDO<sup>11</sup> l'International Health Terminology Standards Development Organisation est une organisation internationale à but non lucratif basée au Danemark. IHTSDO acquiert, détient et gère les droits de la SNOMED CT et d'autres terminologies et des normes connexes. Le but de IHTSDO est de développer, maintenir, promouvoir et permettre l'adoption et l'utilisation correcte de la terminologie de ses produits dans les systèmes de santé, des services et des produits dans le monde entier, et d'entreprendre l'une ou toutes les activités accessoires et propices à la réalisation de l'objectif de l'association pour les avantages des membres. L'IHTSDO vise à améliorer la santé de l'humanité en favorisant le développement et l'utilisation de terminologies cliniques normalisées appropriées, notamment la SNOMED CT, (<http://www.ihtsdo.org/about-ihtsdo/>). L'accent est mis sur l'activation de la mise en œuvre de la sémantique précise des dossiers de santé afin qu'ils

---

<sup>10</sup> [http://www.nlm.nih.gov/research/umls/Snomed/snomed\\_main.html](http://www.nlm.nih.gov/research/umls/Snomed/snomed_main.html)

<sup>11</sup> <http://www.ihtsdo.org/>



soient interoperables. Un support aux membres et titulaires de licence est fourni sur une base globale permettant la mise en commun de ressources. Ce support est à but lucratif.

#### ***I.D.1.c. ADICAP***

**ADICAP**<sup>12</sup> l'Association pour le Développement de l'Informatique en Cytologie et en Anatomie Pathologique est un système de codage développé en France. Gratuit et en français, il est non exhaustif et comporte peu d'entrées pour les anomalies malformatives. Sa mise à jour est irrégulière.

#### ***I.D.1.d. WHO-ICD***

Le **WHO-ICD**<sup>13</sup> est la classification internationale des maladies publiée par l'Organisation Mondiale de la Santé. L'ICD-10 Chapitre XVII (Congenital malformations, deformations and chromosomal abnormalities) comporte peu d'entrées, surtout centrées sur les grands groupes malformatifs.

#### ***I.D.1.e. MeSH, Medical Subject Headings***

Thésaurus biomédical de référence, MeSH<sup>14</sup> est un outil d'indexation, de catalogage et d'interrogation des bases de données de la NLM (National Library of Medicine, Bethesda, USA), notamment de MEDLINE/PubMed. L'objectif du MeSH est de classer hiérarchiquement des termes définis comme uniques (concept), utilisant un vocabulaire contrôlé. Chaque terme MeSH unique est défini par plusieurs champs lui conférant un statut unique, dont un numéro, le MeSH Tree number. Le MeSH est mis à jour annuellement et

---

<sup>12</sup> <http://www.adicap.asso.fr/>

<sup>13</sup> <http://www.dimdi.de/static/en/index.html>

<sup>14</sup> <http://www.nlm.nih.gov/mesh/>

traduit en plusieurs langues. La version 2002 contient 20 742 termes MeSH français pour environ 45 000 synonymes et termes dérivés.<sup>15</sup>

### ***I D.1.f. PathLex***

**PathLex**<sup>16</sup> – Récemment, un groupe de collaboration internationale regroupant l'IHE anatomic pathology<sup>17</sup> (Integrating the Healthcare Enterprise®) et le groupe HL7 anatomic pathology<sup>18</sup> a créé le thésaurus PathLex, qui sera aligné sur le SNOMED CT (Clinical Terms). Il prend en compte les besoins actuels de création de concepts permettant de décrire les phénotypes en anatomie pathologique, en collaboration avec les projets de biobanques liés aux programmes « génome ».

## **I.D.2. Ontologies**

L'ontologie est une notion voisine du thésaurus. En informatique et en science de l'information, une ontologie est une représentation d'un système de connaissance. C'est un ensemble structuré de termes et de concepts fondant le sens d'un champ d'information. L'ontologie constitue en soi un modèle de données représentatif d'un ensemble de concepts dans un domaine, ainsi que les relations entre ces concepts. Employée pour raisonner à propos des objets du domaine concerné, l'objectif premier d'une ontologie est de modéliser un ensemble de connaissances dans un domaine donné<sup>19</sup>. De nombreuses ontologies biomédicales ont été générées (Bodenreider, Mitchell et al. 2005).

Les ontologies biologiques sont fédérées par l'application OLS (Ontology Lookup

---

<sup>15</sup><http://www.caducee.net/Asp/mesh.asp>

<sup>16</sup> <http://www.hl7.org/index.cfm>

<sup>17</sup> [http://wiki.ihe.net/index.php?title=Anatomic\\_Pathology](http://wiki.ihe.net/index.php?title=Anatomic_Pathology)

<sup>18</sup> <http://www.hl7.org/Special/committees/anatomicpath/overview.cfm>

<sup>19</sup> [http://ontology.buffalo.edu/smith/Ontology\\_Course.html](http://ontology.buffalo.edu/smith/Ontology_Course.html)

Services)<sup>20</sup>, développée par l'EMBL<sup>21</sup>(European Molecular Biology Laboratory) qui intègre les différentes ontologies biologiques dans une seule base de données (Cote, Jones et al. 2006; Cote, Jones et al. 2008; Barsnes, Cote et al. 2010; Cote, Reisinger et al. 2010). Les ontologies biologiques sont structurées en un consortium principal appelé OBO pour Open Biomedical Ontologies<sup>22</sup> (Smith, Ashburner et al. 2007). Le consortium OBO coordonne et homogénéise les différentes ontologies en biologie médicale (Bio- ontologies) (Smith, Ashburner et al. 2007). La plus connue est GO<sup>23</sup> (Gene Ontology) qui organise les processus biologiques, les fonctions moléculaires et les composants cellulaires (Ashburner, Ball et al. 2000). Elle est utilisée entre autres par la base de données Ensembl<sup>24</sup>, pour énoncer les fonctions de chaque protéine. Les annotations de Gene Ontology (GO) sont certainement la description la plus compréhensible des fonctions protéiques.

### I.D.3. Ontologies phénotypiques

Il n'existe que très peu d'ontologies phénotypiques, dont les constructions ne font que commencer. Cette construction est un préalable à la mise en place de bases de données phénotypiques exhaustives. Ces ontologies peuvent être plus ou moins fines ou précises. Cette finesse définit leur « **granularité** », selon la profondeur de description.

Chez l'animal, le Jackson Institute<sup>25</sup> (Bar Harbor, Maine) a établi une ontologie des phénotypes de mammifères, mais centrée sur l'étude de la souris ([http://www.informatics.jax.org/searches/MP\\_form.shtml](http://www.informatics.jax.org/searches/MP_form.shtml)). Cette ontologie est utilisée pour

---

<sup>20</sup> <http://www.ebi.ac.uk/ontology-lookup/>

<sup>21</sup> <http://www.embl.de/>

<sup>22</sup> <http://obofoundry.org/>

<sup>23</sup> <http://www.geneontology.org/>

<sup>24</sup> <http://www.ensembl.org/index.html>

<sup>25</sup> <http://www.jax.org/>

la Mouse Phenome Database (MPD)<sup>26</sup> (Paigen and Eppig 2000; Bogue and Grubb 2004; Bogue, Grubb et al. 2007; Grubb, Maddatu et al. 2009). D'autres systèmes sont utilisables chez les animaux, comme PATO (Phenotypic Attribute Trait Ontology) chez le poisson-zèbre (Knowlton, Li et al. 2008).

#### ***I.D.3.a. Vers une ontologie des malformations humaines***

Lorsque nous avons débuté ce travail, il n'existait aucune ontologie des phénotypes humains ou des malformations humaines.

### **I.E. Phénotype, phénomène et phénomique**

Par analogie avec le lien sémantique entre le génome (ensemble des gènes) et le génotype, le phénomène désigne l'ensemble des caractères exprimés (phénotype) par l'organisme humain. Chez l'homme, le phénomène humain est donc constitué par tous les caractères exprimés d'ordre morphologique (macroscopique ou microscopique), biologique (par exemple, la glycémie ou la cholestérolémie), physiologique (par exemple, la tension artérielle) ou neuropsychologique (par exemple les comportements ou les symptômes psychiatriques) (Freimer et Sabatti 2003).

Par analogie avec le couple génome-génomique, la phénomique désigne l'ensemble des technologies déployées pour étudier le phénomène. Il s'agit avant tout de techniques informatiques et de techniques de phénotypage de haut débit. Elle permet la réalisation de projets à large échelle permettant de caractériser des phénotypes de façon standardisée et de corréler les phénotypes observés avec les données de la génomique.

---

<sup>26</sup> <http://phenome.jax.org/>

L'approche phénotypique permet de déterminer des liens entre des combinaisons de traits phénotypiques, d'aider à faire des liens entre variations génétiques et caractères exprimés (corrélations génotype- phénotype). Elle permettra d'identifier des réseaux de maladies, sous-tendus par des réseaux moléculaires ou métaboliques.

### **I.E.1. Bases de données phénotypiques ou phénotypiques**

L'analyse génotype- phénotype détaillée demande l'élaboration de nouveaux outils en informatique pour compiler, trier, intégrer, organiser et gérer les bases de données phénotypiques.

#### ***I.E.1.a. Chez les animaux modèles***

Des bases de données phénotypiques ont été développées chez la souris, en particulier au Jackson Institute (MPD, Mouse Phenome Database, <http://phenome.jax.org/>) (Bogue, Grubb et al. 2007; Masuya, Yoshikawa et al. 2007; Grubb, Maddatu et al. 2009), chez le rat (de la Cruz, Bromberg et al. 2005), chez la levure (Fernandez-Ricaud, Warringer et al. 2007), chez *Caenorhabditis elegans* et *Arabidopsis thaliana* (Kuromori, Wada et al. 2006) au Riken Institute au Japon.

Chez la souris, la base Mouse Phenome Database du Jackson Institute constitue un modèle de ce qui pourrait être fait chez l'homme (Bogue, Grubb et al. 2007; Masuya, Yoshikawa et al. 2007; Grubb, Maddatu et al. 2009). Elle associe des données macroscopiques avec iconographie, des données microscopiques, des données biologiques, des données physiologiques comme la tension artérielle et des données génotypiques relatives aux souches observées. La base de données phénotypiques MausDB a été développée en

Allemagne. C'est une application « open source » pour la gestion des projets de phénotypage à haut débit (Maier, Lengger et al. 2008).

Chez le rat, la base RGD (Rat Genome Database)<sup>27</sup> développe des bases de données phénotypiques<sup>28</sup> (de la Cruz, Bromberg et al. 2005), développées à partir de la Mammalian Phenotype Ontology (Smith, Goldsmith et al. 2005; Smith et Eppig 2009), que l'on peut interroger par l'application PhenoMiner<sup>29</sup>.

### ***1.E.1.b. Chez l'homme***

Chez l'homme, il n'y a pas à ce jour de projet « Phénomène » exhaustif malgré des appels à un « Human Phenome Project »<sup>30</sup> (Freimer et Sabatti 2003), de même ampleur que le « Human Genome Project » des années 1990. Les bases de données phénotypiques (human phenotype data set) décrites plus haut (OMIM, Orphanet, POSSUM) ne peuvent pas être considérées comme des base de données de phénomique.

Deux bases de données de phénomique ont été développées en psychiatrie : une par le Consortium for Neuropsychiatric Phenomics (CNP)<sup>31</sup>, de l'Institut Semel de UCLA (USA) (Bilder, Sabb et al. 2009; Bilder, Sabb et al. 2009; Sabb, Burggren et al. 2009), l'autre sur les troubles bipolaires (Bipolar disorder phenome database, Johns Hopkins School of Medicine) (Potash, Toolan et al. 2007). En Europe, la base de données GenomEUtwin<sup>32</sup> intègre les données phénotypiques et génotypiques de plus de 600 000 paires de jumeaux (Muilu, Peltonen et al. 2007).

---

<sup>27</sup> <http://rgd.mcw.edu/>

<sup>28</sup> <http://rgd.mcw.edu/phenotypes/>

<sup>29</sup> <http://rgd.mcw.edu/wg/news2/11/17---mammary-cancer-data-added-to-phenominer-tool>

<sup>30</sup> <http://www.ncbi.nlm.nih.gov/pubmed/12721547>

<sup>31</sup> <http://www.phenomics.ucla.edu/>

<sup>32</sup> <http://www.genomeutwin.org/>

### ***I.E.1.c. Ponts entre espèces***

Plusieurs équipes se sont intéressées à établir des ponts entre les bases de données phénotypiques. Il s'agit par exemple des projets PhenoHM (Sardana, Vasa et al. 2010), PhenomicDB<sup>33</sup> (Kahraman, Avramov et al. 2005; Groth, Pavlova et al. 2007) and Phenoclustering (Groth, Kalev et al. 2010) du centre de recherche de Bayer Schering. L'équipe d'Anita Burgun (Rennes, France) a développé un système d'annotation automatique des phénotypes de souris à partir de l'index MeSH (Bodenreider et Burgun 2010). Ils ont également bâti une carte terminologique entre la Mammalian Phenotype Ontology (MPO) et OMIM par l'intermédiaire de système UMLS (Unified Medical Language System) (Burgun, Mougin et al. 2009).

### ***I.E.1.d. Représentation des phénotypes pour l'analyse à grande échelle***

Ces ontologies phénotypiques permettent une annotation à grande échelle de phénotypes analysés d'une façon structurée et cohérente. Un modèle de cette annotation est donné par le projet « The Phenotype Quality Ontology » (PATO), anciennement nommé « Phenotype Attribute and Trait Ontology » (Marquet, Mosser et al. 2007; Knowlton, Li et al. 2008; Gkoutos, Mungall et al. 2009). PATO est en voie de devenir la nouvelle norme pour annoter les phénotypes analysés d'une façon structurée et cohérente dans diverses espèces biologiques. Cet outil est appuyé par le National Center for Biomedical Ontology<sup>34</sup>, et il a été évalué par le Gene Ontology Consortium (Knowlton, Li et al. 2008).

---

<sup>33</sup> <http://www.phenomicdb.de/>

<sup>34</sup> <http://www.obofoundry.org/>

***I.E.1.e. Phénomique et phénotypes malformatifs***

La dysmorphogenèse humaine constitue un champ idéal pour débiter un projet de phénomique humaine. En effet, il s'agit d'un champ clos constitué de combinaisons variables d'environ 1000 éléments phénotypiques, appelés « anomalies développementales ».



## **Chapitre II : Objectifs de notre travail**

Les objectifs de notre travail étaient multiples :

- évaluer la faisabilité des approches phénoquiques en pathologie humaine en appliquant ces techniques à l'étude des malformations humaines
- élaborer un thésaurus et une ontologie des malformations humaines dans le cadre plus large d'une ontologie des caractères phénotypiques humains
- stocker dans une base de données informatique les constatations des autopsies fœtales réalisées au CHU Sainte-Justine et effectuées dans un cadre clinique d'études malformatives.
- déterminer la fréquence de chacune des anomalies développementales observées dans la série d'autopsies fœtales du CHU Sainte-Justine
- définir de nouvelles combinaisons malformatives
- fournir une aide informatisée au diagnostic des combinaisons malformatives humaines
- élaborer une liste de gènes candidats en fonction des agrégations phénotypiques observées
- élaborer une carte en réseau des malformations humaines

## **Chapitre III : Matériel et méthodes**

### **III.A. Généralités sur l'examen fœtopathologique**

La fœtopathologie est une sous-spécialité de la pathologie qui porte sur l'étude morphologique des maladies fœtales, et en particulier des malformations. Si une interruption de la grossesse est pratiquée ou si le décès du fœtus survient spontanément, le pathologiste du développement confirme ou apporte des éléments complémentaires de diagnostic visant à faire progresser les connaissances sur la pathologie identifiée. Cette étude est nécessaire pour répondre aux questions sur les causes de décès, les anomalies de développement et les probabilités quant aux risques d'une présentation identique ou semblable lors de grossesses ultérieures.

Cette analyse suit un protocole bien établi, incluant l'analyse du placenta. Les résultats obtenus doivent faire l'objet d'une concertation dans un groupe multidisciplinaire composé de l'obstétricien, du généticien, du néonatalogiste, du radiologiste et du pathologiste. Cette sous-spécialité de la pathologie englobe également toute recherche sur le matériel d'avortement spontané, précoce ou tardif, ainsi que le matériel provenant d'interruption thérapeutique. L'analyse de ce matériel porte sur les anomalies génétiques, les troubles de l'organogenèse, les maladies métaboliques, les infections, les troubles de maturation fœtale et les maladies maternelles entraînant des effets néfastes sur le développement du fœtus ou du nouveau-né. Pour compléter cette recherche, un examen approfondi du placenta s'avère essentiel pour la compréhension de certaines pathologies. L'ensemble de ces analyses se révèle important pour l'information donnée aux parents et le conseil génétique qui en découle est une aide précieuse pour la prise en charge des grossesses suivantes de la mère et de son conjoint, de même que pour la famille proche. Ainsi, le pathologiste doit faire partie intégrante d'une équipe multidisciplinaire qui s'intéresse à la femme parturiente et à son fœtus.

### **III.B. L'autopsie fœtale au CHU Sainte- Justine**

L'autopsie fœtale au CHU Sainte-Justine suit un protocole bien établi. L'ensemble de ces analyses se révèle important pour l'information donnée aux parents et le conseil génétique qui en découle est une aide précieuse pour la prise en charge de grossesses subséquentes.

#### **III.B.1. Photographies**

Des **photographies** générales ou détaillées du fœtus sont produites comme base de données. Les photographies sont faites par un photographe médical, avec une prise de face et de profil; d'autres prises sont effectuées lorsque jugées nécessaires. Les photos sont stockées dans la base de données clinique.

#### **III.B.2. Radiographies**

Des **radiographies** du fœtus sont pratiquées dans la presque totalité des cas où une malformation est connue ou soupçonnée afin de vérifier l'aspect du profil osseux, le développement adéquat de l'ossature, incluant une anomalie de configuration ou une ossification précipitée. L'âge osseux est également évalué par rapport à l'âge gestationnel présumé pour évaluer la possibilité d'un retard de croissance intra-utérin. Une radiographie nous permet de visualiser la présence ou non d'éléments inattendus comme de l'air formant un pneumothorax provoqué à la suite d'une ventilation artificielle chez le prématuré par exemple. Toutes les radiographies fœtales sont évaluées par un des radiologistes pédiatriques du CHU Sainte-Justine, et notre étude de cas est basée sur les rapports diagnostiques au dossier

### **III.B.3. Culture de fibroblastes**

Une **culture de fibroblastes** est effectuée pour l'obtention d'un caryotype ou l'éventuelle recherche d'une maladie métabolique lorsque indiqué. Même en cas de mort in utero, cette approche peut être envisagée, en prélevant de façon stérile un fragment de peau préalablement nettoyé à l'alcool 70 % à l'aide d'un bistouri et d'une pince stériles.

### **III.B.4. Examen morphologique**

L'examen morphologique du fœtus est essentiel au diagnostic d'un bon nombre de lésions génétiques à révélation prénatale. Cet examen porte sur l'étude des organes in situ, quant à leur position, leurs relations et leur développement. L'autopsie se fait généralement par l'assistant pathologiste sous supervision du pathologiste pédiatrique.

### **III.B.5. Examen histologique**

L'examen histologique revêt également un caractère particulier. Chaque organe est évalué en fonction de l'âge gestationnel pour repérer toute asynchronie ou défaut de formation. Comme les tissus fœtaux sont en développement continu, le pathologiste doit toujours se référer non seulement à l'histologie conventionnelle de chaque tissu mais également tenir compte de l'âge gestationnel. Cette analyse histologique pourrait permettre de répondre aux questions soulevées par les parents en vue d'une grossesse ultérieure. Cette étape finale de l'autopsie fœtale est grandement facilitée et complémentée par une concertation multidisciplinaire avec l'obstétricien, le néonatalogiste, le généticien et radiologiste.

### **III.B.6. Examens complémentaires**

Un morceau de foie ou un segment de 2-3 cm du cordon (à la base du placenta) est congelé à -80 °C pour examen éventuel en biologie moléculaire eg. PCR ou RT-PCR (PCR avec transcriptase inverse), Q-PCR (PCR quantitatif), et hybridation in situ et études par micro-puces.

### **III.B.7. Examens de microbiologie et de virologie**

Ces études s'ajoutent aux analyses de routine, et parfois aux techniques de biologie moléculaire comme par exemple une Q-PCR pour recherche d'une infection à Parvovirus B-19.

### **III.B.8. Microscopie électronique**

Ces études sont particulièrement utiles pour la recherche de matériel de surcharge, dans un contexte de maladie métabolique et pour la recherche d'inclusions virales dans le cas d'infection.

### **III.B.9. Softpath**

Softpath est un logiciel de gestion des laboratoires dédié à la pathologie. Il s'agit du système informatique de gestion des données du département pathologie du CHU Sainte Justine depuis l'an 2000. Grâce à ses outils flexibles, Softpath permet aux médecins pathologistes de créer facilement un rapport standardisé d'autopsie. Ce rapport comprend le nom du pathologiste; le nom de patient; le numéro de dossier médical du patient. Ainsi il est possible d'obtenir par Softpath les comptes-rendus macroscopiques et microscopiques

des spécimens chirurgicaux et des autopsies, de biologie moléculaires et de cytologie, et d'accéder aux antécédents des patients. Il est également possible d'utiliser Softpath pour faire des recherches de cas à partir de code diagnostics (SNOMED) ou par mot clé. Softpath est également utilisé par les assistants pathologistes et les techniciens qui y rentrent les données cliniques et les techniques effectuées pour chaque cas.

### III.C. Procédures de fonctionnement de la base de données MDB

Nous (moi-même-Farideh Arbabzadeh, Dr Jean-Christophe Fournet, et M.Philippe Sawicki) avons mis au point une base de données des malformations fœtales (MDB) grâce au logiciel Microsoft Access et au pseudo- langage informatique MySQL.

#### III.C.1. Access

Access<sup>35</sup> est un système de gestion de base de données relationnelle qui est compatible avec les requêtes SQL ; il dispose d'une interface graphique pour saisir les requêtes.

Access est caractérisé par des enregistrements et des champs. Dans le cas d'un carnet d'adresse, l'**enregistrement** représente les coordonnées de la personne et le **champ** représente la case à remplir. La création d'une base de données implique donc au préalable la création d'une structure en fonction des renseignements souhaités.

Les onglets d'Access se subdivisent en six parties :

- Une **table**, reprend les informations brutes (non traitées): les enregistrements et les champs.
- Une **requête**, permet diverses mises en forme des tables de trier, en plus de filtrer et d'afficher le nom des données.

---

<sup>35</sup> <http://access.developpez.com/>



- **Un Formulaire**, affiche les données des tables et les requêtes d'une manière convenable.
- **Un état**, permet d'imprimer des données de tables et de requêtes avec, si nécessaire, des regroupements et des calculs.
- Les **macros**, permettent d'exécuter des fonctions à partir de requêtes ou de formulaires.
- **Page**, permet de créer des pages Internet pour accéder à la base de données

### III.C.2. Une base de données en langage MYSQL

Le langage MySQL<sup>36</sup> (**Structured Query Language SQL**, ou **langage structuré de requêtes**) est un pseudo-langage informatique (de type requête) standard et normalisé. Il sert à construire des bases de données relationnelles. Une **base de données relationnelle** est une base de données structurée qui permet de stocker et d'organiser une grande quantité d'information. Les SGBD (Système de Gestion de Base de Données) permettent de naviguer dans ces données, d'extraire et de mettre à jour les informations voulues au moyen d'une requête (une requête correspond à l'interrogation d'une base pour en récupérer une certaine partie des données). Le langage MySQL permet de piloter un serveur de base de données SQL très rapide, multi-utilisateur et robuste. Le serveur MySQL est destiné aux missions stratégiques et aux systèmes de production à forte charge, ainsi qu'à l'intégration dans des logiciels déployés à grande échelle. MySQL est une marque déposée de MySQL AB.

---

<sup>36</sup> [www.mysql.com/](http://www.mysql.com/)

### **III.C.3. Le type de données recueillies**

- Dans une base locale au format Access :
  - Numéro de dossier HSJ
  - Numéro d'autopsie HSJ
  - Numéro d'étude
- Dans une base MySQL en ligne totalement codée :
  - Numéro d'étude
  - Anomalies du développement constatées
  - Diagnostic

### **III.C.4. Le niveau d'identification des données et échantillons (codées, anonymisées)**

Deux bases sont créées :

- Une base de données locale au format Access ne stockant que le code d'étude locale et des informations sur un fœtus donné :
  - Numéro d'autopsie
  - Numéro de dossier
- Une base de données codée, basée sur le Web (web-based), totalement codée, sans aucune donnée nominative ou moyen d'identification.

### **III.C.5. Le lieu de conservation des données**

- La base Access verrouillée est située dans un dossier 'Malformation Database', dans le dossier « Pathologie », dans le dossier « public hsj » sur

le disque dur « Snetapp02 » du CSIT (Centre de Support Informatique) du CHU Sainte-Justine. Ce dossier 'Malformation Database' est d'accès restreint à deux personnes: Dr Jean-Christophe Fournet et moi-même, Farideh Arbabzadeh, en fonction de leur identification (ID) et de leur mot de passe (PW) par le CSIT du CHU Saine- Justine.

- La base de données MySQL (MDB, Malformation DataBase) codées sur un serveur MySQL distant (servage.com).

Aucune autre personne ne peut accéder aux données nominatives (base Access local).

Les données de la base MySQL basée sur le Web sont accessible librement sur le Web.

Rien ne permet leur identification.

### **III.C.6. Convention régissant l'utilisation des données du matériel biologique**

Une convention qui régira l'utilisation des données du matériel biologique par les chercheurs est établie.<sup>37</sup> Par cette convention, les chercheurs s'engagent à conserver la confidentialité des données du matériel biologique, à les utiliser à des fin pour lesquelles la base de données a été établie et pour lesquelles le consentement a été obtenu. Les chercheurs s'engagent à ne pas transmettre les données à des personnes non autorisées.

### **III.C.7. L'approbation du comité d'éthique de la recherche (CÉR)**

Notre projet de recherche est une étude rétrospective de dossiers anatomo-pathologique et il n'y a pas d'analyse de tissus de fœtus pour cette étude.

---

<sup>37</sup> ASSOCIATION MÉDICALE MONDIALE, *Declaration on Ethical considerations Regarding Health Databases*, Washington 2002, art. 23.

Une lettre d'accord de la DSP<sup>38</sup> (Directeur des Services Professionnels) a été obtenue pour accéder à ces dossiers :

- L'accès au dossier médical et de la divulgation de l'information
- L'obtention de consentement des parents à l'autopsie
- L'accès au dossier médical de la mère

Les banques de données du matériel biologique sont soumises à la *politique d'examen continu du projet de recherche en cours* du CÉR au CHU Sainte-Justine.

La base de données MDB sera utilisée conformément aux objectifs pour lesquels elle a été créée avec l'autorisation du CÉR<sup>39</sup>.

---

<sup>38</sup> [http://dsp.aqesss.qc.ca/997/Site\\_DSP.aqesss](http://dsp.aqesss.qc.ca/997/Site_DSP.aqesss)

<sup>39</sup> Code civil du Québec, art. 37 et art 5; Loi sur la protection des renseignements personnels dans le secteur privé et dans les organismes publics

## **Chapitre IV : Les étapes de notre travail**

La première étape de la mise en place d'une approche phénotypique des malformations humaines est la construction d'une banque de données des anomalies développementales humaines, qui servira de base aux traitements automatisés ultérieurs. La conception et la mise en charge de cette base de données a constitué mon travail de stage et le sujet de mon mémoire de maîtrise.

Notre travail s'est bâti sur 4 étapes successives :

**IV-A. Élaboration d'une base de données des malformations humaines**

**IV-B. Élaboration d'une ontologie des malformations fœtales**

**IV-C. Saisie des données phénotypiques des autopsies de fœtus malformés**

**IV-D. Publication de la base de données sur le réseau Internet**

## **IV.A. Élaboration d'une base de données des malformations humaines**

Dans un premier temps, nous avons construit une base de données, nous permettant de construire une ontologie des termes utilisés en pathologie malformative, de stocker et de traiter nos données phénotypiques. Cette base de données a été construite en collaboration avec M. Philippe Sawicki, stagiaire de l'École Polytechnique, en langage MYSQL.

## **IV.B. Élaboration d'une ontologie des malformations fœtales**

La phase préparatoire à la mise en place d'une ontologie au format OBO dans un domaine donné est la constitution d'un thésaurus exhaustif dans ce domaine. L'ontologie est ensuite créée avec un logiciel de production d'ontologie (comme OBO-edit).

Au cours de ce travail, j'ai élaboré trois thésaurus sur les malformations humaines, stockés

dans une base de données MySQL accessibles sur le web à l'adresse <http://www.malformations.org> :

- un thésaurus des anomalies développementales humaines
- un thésaurus des combinaisons malformatives
- un thésaurus des maladies malformatives

#### IV.B.1. Construction d'un thésaurus des anomalies développementales humaines

Comme première étape de ce travail, nous avons réalisé un thésaurus des anomalies développementales humaines. Ce thésaurus est une dénomination et une description de toutes les anomalies du développement fœtal observées dans la série d'autopsies fœtales effectuées pendant 6 ans, de 2001 à 2006, au CHU Sainte-Justine. Il compte aujourd'hui 751 entrées, et est encore en cours d'extension. Ce thésaurus a la vocation d'être exhaustif et sera complété au cours des prochaines années.

Pour chaque anomalie décrite dans le thésaurus, nous avons entré dans notre base de données sa **définition**, ses **synonymes**, sa **localisation**, son **type** et **l'anomalie du développement parent**, de même que leurs **combinaisons malformatives** associées.

Un exemple de fiche « anomalie » est fourni à l'adresse suivante :

<http://www.malformations.org/thesaurus/view/35>

Sur cette fiche « anomalie », on trouve :

- Le numéro de l'anomalie (DA : #)
  - Ce numéro est très important car c'est lui qui désigne l'entité décrite.
  - Il ne change pas même si le nom peut changer avec le temps.
- Le nom de l'anomalie

- A chaque numéro est associé un nom qui peut changer ou être remplacé par un synonyme.
- Les synonymes
  - La recherche des synonymes est très importante car une même entité peut avoir plusieurs noms.
- La définition de l'anomalie
  - Les procédures d'ontologie telles que définies par le consortium OBO requièrent que la définition de chaque entité décrite soit explicitée
  - La définition peut être variable dans le temps
  - Cette définition a été sélectionnée et décrite parmi les dizaines descriptions d'anomalies fœtales dans les littératures médicales
  - le mode écriture présente la page sous une forme qui permet de la modifier
  - la définition est modifiable par des personnes ou des organismes qui auront accès aux données de la base de données MDB
- Catégorie
  - Chaque anomalie est classée selon sa catégorie supposée (malformation, dysgénésie, agénésie).
- Code de topologie
  - Il indique la topologie de cette anomalie.
- Code de système
  - Il indique dans quel système prend place l'anomalie (par exemple, système respiratoire ou cardiovasculaire).
- Code « parent » d'anomalie (anomalies du développement de parent)
  - La notion de code « parent » est très importante. Il permet d'établir une



hiérarchie dans les anomalies constatées; par exemple, le code parent de « *Polydactylie post-axiale* » est « *Polydactylie* ». Le code parent de « *Polydactylie* » est « *Anomalie des doigts* ».

- Des liens sont établis avec d'autres sites web comme :
  - [hupath.com](http://hupath.com/)<sup>40</sup>
  - HPO (Human Phenotype Ontology)<sup>41</sup>
  - Wikipedia<sup>42</sup>
  - OMIM<sup>43</sup>

#### **IV.B.2. Construction d'un thésaurus des combinaisons malformatives humaines**

Parallèlement, nous avons commencé un thésaurus des combinaisons malformatives dans lesquelles ces anomalies peuvent s'associer. Ce thésaurus est accessible à l'adresse :

<http://www.malformations.org/combinations>

La plupart d'entre elles ont été listées dans la base de données OMIM, cependant, cette base de données porte plus sur les maladies mendéliennes et leurs synopsis cliniques ( voir annexes 3 et 4) énumèrent des anomalies développementales selon un langage non contrôlé.

A titre d'exemple, quarante trois combinaisons malformatives observées dans notre série ont ainsi été stockées. Elles sont associées à un ensemble d'anomalies développementales préalablement annotées et numérotées dans le cadre de notre ontologie.

---

<sup>40</sup> <http://www.hupath.com/>

<sup>41</sup> [http://www.human-phenotype-ontology.org/index.php/hpo\\_docu.html](http://www.human-phenotype-ontology.org/index.php/hpo_docu.html)

<sup>42</sup> <http://fr.wikipedia.org/>

<sup>43</sup> <http://www.ncbi.nlm.nih.gov/omim/>

Pour chaque combinaison malformative, les données suivantes sont stockées :

- Numéro de la combinaison (SYD:#)
- Nom de la combinaison
- Synonymes
- Définition
- Description
  - Il s'agit de la liste des anomalies qui constituent la combinaison malformative.
  - Contrairement à OMIM, il s'agit d'une liste d'entités numérotées et dénommées selon un langage contrôlé.
- Cas de l'étude présentant cette maladie malformative
- Références
  - Il s'agit des codes attribués par les autres ontologies (OMIM, SNOMED, MeSH) ou bases de données (OMIM, humpath.com, Orphanet, e-medecine).
  - Ces références permettent d'établir des liens hypertexte avec ces bases de données.
- Catégorie de cette combinaison

Il est très difficile de catégoriser les combinaisons malformatives. On peut distinguer néanmoins quatre groupes principaux

- les **défauts des champs polytopiques** (Polytopic Field Defects);  
ex: les anomalies du situs comme le lévo-isomérisme
- les **syndromes malformatifs**: Ensemble d'anomalies dérivant toutes de la même cause et ne correspondant ni à une **séquence** ni à une

**association;** ex : syndrome de Goldenhar

- les **associations**: Survenue non fortuite d'au moins deux malformations non reconnues comme séquence ou syndrome; ex: association VACTERL
- les **séquences**: Ensemble d'anomalies qui sont toutes la conséquence en cascade d'une seule anomalie préalable ou d'un facteur mécanique ; ex : séquence oligohydramnios (dite encore de Potter), séquence de Pierre Robin, séquence de cordon ombilical court

Plusieurs groupements d'anomalies observés dans notre étude ne correspondent pas à des entités décrites.

#### **IV.B.3. Construction d'un thésaurus des maladies malformatives humaines**

Nous avons également créé un thésaurus des maladies malformatives,<sup>44</sup> qui sont l'ensemble des anomalies, éventuellement groupées en combinaisons, causées par une étiologie définie : une anomalie chromosomique, une anomalie monogénique ou une infection. Ce thésaurus est accessible à l'adresse :

[http://www.malformations.org/malformative\\_diseases](http://www.malformations.org/malformative_diseases)

Pour chaque maladie malformative, la base MDB stocke les données suivantes :

- Numéro (DIS:#)
- Nom de la maladie
- Synonymes

---

<sup>44</sup> [http://www.malformations.org/malformative\\_diseases](http://www.malformations.org/malformative_diseases)

- Définition
- Catégorie selon l'étiologie
- Anomalies développementales observées (numérotées et contrôlées)
- Combinaisons malformatives observées
- Cas de l'étude présentant cette maladie malformative
- Cause de la maladie
- Codes des autres ontologies
- Liens avec les autres bases de données

#### **IV.B.4. Bilan**

En juillet 2009, le nombre d'entrées disponibles pour notre base de données était de :

- Anomalies développementales (751 éléments)
- Combinaisons malformatives (43 éléments)
- Maladies malformatives (41 éléments)

#### **IV.C. Saisie des données phénotypiques des autopsies de fœtus malformés**

Afin de diffuser et acquérir des connaissances sur les combinaisons polymalformatives observées en diagnostic anténatal ainsi que l'analyse phénotypique des malformations, nous avons décidé d'étudier rétrospectivement la série des autopsies de fœtus du CHU Sainte-Justine entre 2001 et 2006, et d'élaborer une base de connaissances sur les malformations humaines (<http://www.malformations.org>). Les données de 543 autopsies présentant une ou des malformations ont été saisies afin de la mettre en charge et d'en tester les fonctions principales (l'étude pilote). Ces données sont accessibles à l'adresse : <http://www.malformations.org/case>

### **IV.C.1. Élaboration d'une fiche de saisie des données phénotypiques**

Pour chaque cas étudié <sup>45</sup> les données suivantes ont été stockées :

- Numéro du cas local (HSJ#)
- Numéro du cas global (MDB#)
- Sexe fœtal
- Renseignements cliniques
- Commentaires
- Liste des anomalies observées contrôlées par le thésaurus
- Diagnostic

### **IV.C.2. Étude rétrospective de la série d'autopsies des fœtus du CHU Sainte-Justine entre 2001 et 2006# Saisie des données fœtopathologiques dans la base de données MDB**

Nous avons réalisé une étude rétrospective de la série des autopsies des fœtus du CHU Sainte-Justine entre 2001 et 2006. A partir des rapports d'autopsies, les anomalies de chaque fœtus étudié ont été stockées dans la base de données MDB selon le thésaurus que nous avons préalablement établi.

#### ***IV.C.2.a. Population cible***

Notre population cible, aussi appelée population à l'étude, fut constituée des fœtus porteurs d'au moins une anomalie développementale et caractérisés par une étude autopsique dans le département de pathologie du CHU Sainte-Justine entre 2001 et 2006. Il s'agissait principalement de fœtus ayant fait l'objet d'un dépistage anténatal échographique de

---

<sup>45</sup> <http://www.malformations.org/case>

malformations avec interruption thérapeutique de grossesse entre 18 et 22 semaines gestationnelles (fœtus moins de 500 gr) ou de fœtus malformés décédés dans la période périnatale. Ces anomalies malformatives pouvaient être dues à des anomalies génétiques, des anomalies de l'organogenèse, des maladies métaboliques, des disruptions ou des infections.

#### ***IV.C.2.b. Collecte des données***

Les numéros d'autopsie et de dossier ont été retracés à partir du logiciel Softpath (base de donnée de département pathologie du CHU Sainte-Justine) en important les données vers Microsoft Excel. Un numéro d'étude par l'intermédiaire d'un fichier à accès protégé a été alloué à chaque cas.

Les paramètres suivants ont été stockés :

- Numéro de dossier du CHU Sainte-Justine
- Numéro d'autopsie
- Numéro d'étude (HSJ)

Un total de 543 fœtus avec anomalie(s) développementale(s) a été étudié entre 2001-2006 par le département de pathologie du CHU Saint-Justine et est inclus dans cette étude. Un numéro d'étude local (HSJ#) et un numéro d'étude global (MDB#) ont été alloués pour chaque fœtus étudié. Comme le CHU Sainte-Justine est à ce jour le seul centre participant, le numéro local **HSJ** et le numéro global **MDB** sont identiques.

Les quatre rapports de chaque dossier (macroscopie, microscopie, neuropathologie, conclusions finales) ainsi que les photos stockées dans le logiciel Softpath ont été revus. Les données pertinentes ont été extraites et intégrées manuellement dans la base de données MDB.

#### **IV.D. Publication de la base de données sur le réseau Internet .**

La base de données anonymisée au format MySQL est publiée en temps réel sur le réseau internet. <http://www.malformations.org/>.

## **Chapitre V : Discussion et perspectives**



Pour notre travail de maîtrise, nous avons donc créé une base de données permettant de stocker des phénotypes observés dans une série autopsique de fœtus malformés. Au cours de cette étude-pilote, 543 phénotypes ont été stockés dans notre base de données MDB accessible à l'adresse web <http://www.malformations.org>. Pour la partie « Thésaurus », les entrées disponibles en juillet 2009 étaient 751 anomalies développementales (751 éléments), 43 combinaisons malformatives, 41 maladies malformatives.

## **V.A. Automatisation de la collecte de données**

Au cours de ce travail, les données ont été collectées de façon manuelle pour cette étude-pilote de preuve de concepts et de faisabilité. Or les technologies phénomiques reposent sur des acquisitions très rapides (haut débit) et automatisées d'un très grand nombre de données. A ce titre, cette étude-pilote a permis de bien analyser les différentes étapes de la collecte et les différents verrous ou goulots d'étranglement méthodologiques.

En collaboration avec le centre de recherche européen de la compagnie Xerox<sup>46</sup> (le XRCE - basé en France à Grenoble), notre équipe réfléchit actuellement à des méthodes automatisées de lecture des dossiers et des rapports médicaux, de sélection des données pertinentes et d'intégration automatique dans la base de données.

## **V.B. Les usages des bases de données phénotypiques et phénomiques**

Notre étude-pilote sur 543 fœtus malformés permet d'entrevoir les différents usages d'une base de données phénotypiques dans ce domaine.

---

<sup>46</sup> [www.silicon.fr/xerox-ouvre-les-portes](http://www.silicon.fr/xerox-ouvre-les-portes)

### **V.B.1. Importance des statistiques et des données quantitatives**

La base de données MDB est la seule base de données à produire des données statistiques (voir annexe 5- page xxi). Elle permet de calculer des fréquences d'observation dans des syndromes donnés. Ces données seront très utiles dans l'aide au diagnostic ainsi que dans les études épidémiologiques.

### **V.B.2. Les corrélations génotype- phénotype**

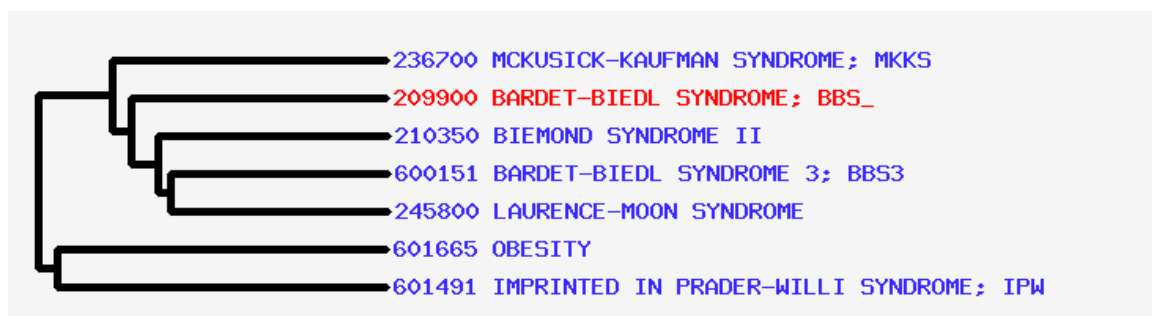
La description de dizaines de milliers de variants génotypiques (mutations, SNPs- Single Nucleotide Polymorphisms, CNVs- variations du nombre de copies- Copy Number Variations) impose l'accès aux chercheurs d'outils performants de stockage des données phénotypiques. Ce stockage permettra la mise en évidence de liens statistiques entre une variation génétique et un ou plusieurs traits phénotypiques. Si de nombreuses mutations de gènes ont déjà été mises en évidence dans les syndromes malformatifs, de nombreux CNVs ont été récemment associés avec des phénotypes malformatifs (Kriek, White et al. 2006; Shinawi and Cheung 2008). Notre base de données MDB permettra de stocker ces données.

### **V.B.3. Distance phénotypique**

La similarité ou la différence entre les syndromes malformatifs mesure la « distance phénotypique ». La mesure de cette distance permet de définir les concepts liés de « proximité phénotypique » ou d' « éloignement phénotypique » (voir annexe 3- pages xvi et xvii).

Cette notion a été appliquée par les auteurs de l'application MimMiner<sup>47</sup>, qui mesure la distance entre les phénotypes publiés dans OMIM selon les nombre de caractères partagées dans le synopsis clinique de OMIM (van Driel, Bruggeman et al. 2006).

MimMiner établit un dendrogramme des phénotypes observés en fonction de la proximité des phénotypes observés. En fonction du nombre de traits phénotypiques partagés, l'application MimMiner mesure la distance phénotypique entre deux syndromes donnés.



**Figure 2 :** Dendrogramme produit par la base MimMiner par l'entrée « syndrome de Bardet-Biedl ». MimMiner crible les synopsis cliniques d'OMIM pour trouver les syndromes malformatifs (entrées OMIM) les plus proches. A noter que ces syndromes retrouvés et la proximité des phénotypes sont des ciliopathies, des anomalies moléculaires de la régulation du cil primaire (voir annexe 3).

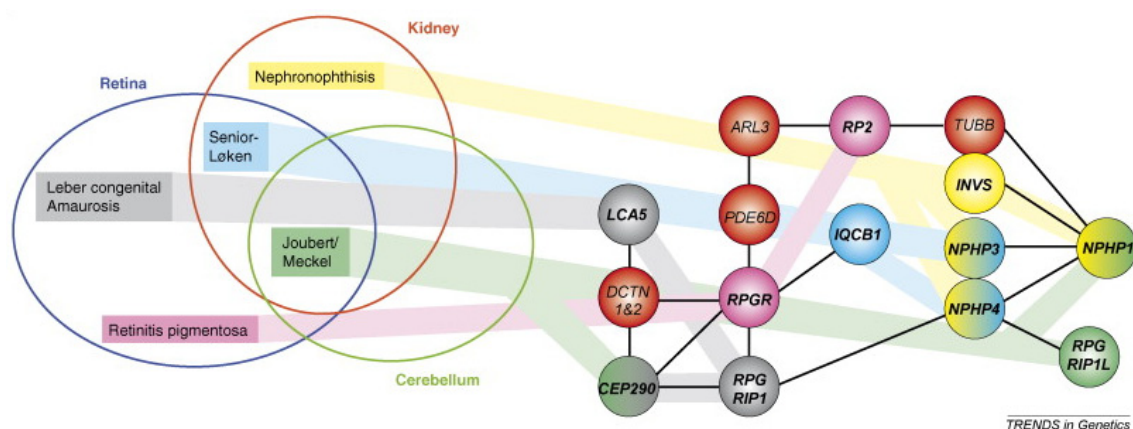
Dans une autre approche, Oti. et al. ont développé la notion de « distance phénotypique ». En étudiant conjointement les bases de données OMIM, Orphanet et POSSUM, ils ont pu mesurer les distances entre les syndromes et définir des familles de syndrome (*syndrome families*) ou agrégats phénotypiques (*phenotype clusters*), comme par exemple la maladie d'Ehlers-Danlos ou la maladie de Charcot-Marie-Tooth (Oti, Huynen et al. 2009). Ils ont montré une corrélation entre les agrégats phénotypiques et la proximité des annotations GO (gene ontology) des gènes qui causent les maladies agrégées dans ces familles (Oti, Huynen et al. 2009).

<sup>47</sup><http://www.cmbi.ru.nl/MimMiner/help.html#About>

#### **V.B.4. Agrégation phénotypique et familles de maladies**

Des maladies peuvent être classées dans une même « famille de maladies » si les phénotypes produits sont voisins, et produisent des « chevauchements » phénotypiques (Brunner and van Driel 2004) (voir annexe 4- pages xviii, xix, xx ). Ces chevauchements phénotypiques et les difficultés phénotypiques qu'ils occasionnent sont à l'origine du vieux débats entre les diviseurs et les fusionneurs (the « splitter and lumpers debate »), décrit par Victor McKusick lui-même en 1969 (McKusick 1969). Ce débat est maintenant arbitré par la génétique moléculaire qui a permis de résoudre plusieurs problématiques de chevauchement phénotypique et de créer des familles moléculaires de maladies.

L'archétype de ces familles moléculaires de maladies est le groupe des « ciliopathies » (Badano, Mitsuma et al. 2006; Quinlan, Tobin et al. 2008; Tobin and Beales 2009). En effet, le groupe de maladies constitué par le syndrome de Meckel-Gruber, le syndrome de Joubert, la néphronophtisie, l'amaurose congénitale de Leber et le syndrome de Senior-Loken est le siège de phénotypes chevauchant (overlapping phenotypes) associant dysplasie multikystique rénale (MKD), dégénérescence rétinienne, polydactylie et malformations cérébrales. Ce groupe a pu être unifié lorsqu'il est apparu que les gènes mutés dans ces maladies mendéliennes codaient pour des protéines régulant le fonctionnement du cil primaire (« ciliopathies ») (Badano, Mitsuma et al. 2006).



**Figure 3 :** La famille des « ciliopathies ». Ces phénotypes chevauchants sont causés par des anomalies de protéines interagissant entre elles. *D'après* Oti, M., M. A. Huynen, et al. (2008). "Phenome connections." Trends Genet. 24(3):103-106.

Cet exemple princeps montre qu'un chevauchement phénotypique entre des maladies évoque une physiopathogénie commune impliquant des protéines aux fonctions voisines (Brunner et van Driel 2004; Oti, Huynen et al. 2008) ou interagissant dans un même complexe moléculaire (Oti, Snel et al. 2006). La parenté moléculaire des maladies dessine ainsi un réseau de maladies (disease network) en fonction de voies de signalisation et d'interactions moléculaires (Oti, Snel et al. 2006; Goh, Cusick et al. 2007; Lage, Karlberg et al. 2007; Wu, Liu et al. 2009).

Ainsi, dans l'étude de Goh et al., les gènes associés à des maladies voisines montraient une plus forte probabilité d'interaction de leurs produits protéiques et une plus forte similarité de leurs profils tissulaires d'expression, soutenant l'existence de « modules » fonctionnels spécifiques de maladies (Goh, Cusick et al. 2007). De la même façon, van Driel et al. ont montré que la proximité des phénotypes était liée à plusieurs paramètres fonctionnels des gènes, en particulier la similarité de la séquence protéique, les motifs protéiques, l'annotation fonctionnelle selon GO (Gene Ontology) et les interactions protéine- protéine directes (van Driel, Bruggeman et al. 2006). Il est également possible de lier les maladies

génétiques entre elles (genetic disease network) par les annotations Gene Ontology des gènes mutés (Perez-Iratxeta, Bork et al. 2002) ou par les interactions protéine- protéine directes de leurs protéines produites (van Driel, Bruggeman et al. 2006).

Ces familles moléculaires ont été appelées « modules », et ont créé le concept de « modularité » des maladies génétiques (Oti and Brunner 2007; Jiang, Liu et al. 2008).

### **V.B.5. Proximité phénotypique - Proximité protéique**

Les annotations de Gene Ontology (GO) constituent la description la plus compréhensible des fonctions protéiques. Ces fonctions permettent de mesurer la similarité fonctionnelle des produits protéiques de ces gènes. Il existe cependant d'autres façons de mesurer la similarité de produits protéiques : des voies biochimiques communes (Kanehisa M, 2004), des domaines protéiques communs (Hunter S, 2009), des interactions protéines- protéines.

### **V.B.6. Agrégation phénotypique**

Les distances entre les combinaisons phénotypiques permettent de définir des agrégats phénotypiques (*phenotypic clusters*) définis par la proximité des phénotypes entre eux (Johnson, Barton et al. 1981; Jaqaman, Dorn et al. 2007). Ainsi, les dendrogrammes précédents définissent des agrégations phénotypiques (*phenotypic clustering*) de ce type.

### **V.B.7. Réseau phénotypique**

Il est possible d'établir des réseaux entre les maladies selon les anomalies partagées, ou entre les anomalies selon les maladies partagées. Ainsi, Lazlo Barabasi a construit récemment un schéma globale des maladies humaines selon la fréquence des associations observées entre ces entités (Barabasi 2007; Goh, Cusick et al. 2007; Midic, Oldfield et al.

2009). Il a ainsi créé une approche globale des maladies humaines vues comme un tout (diseasome<sup>48</sup>) (Barabasi 2007; Midic, Oldfield et al. 2009), et créé une nouvelle approche technologique la « diseasomics » (voir annexe 6- page xxii). Pour étudier les agrégations phénotypiques et les associations (co-occurrence) de maladies, il est très utile d'étudier de grandes quantités de données tirées du codage hospitalier (PMSI) (Rzhetsky, Wajngurt et al. 2007).

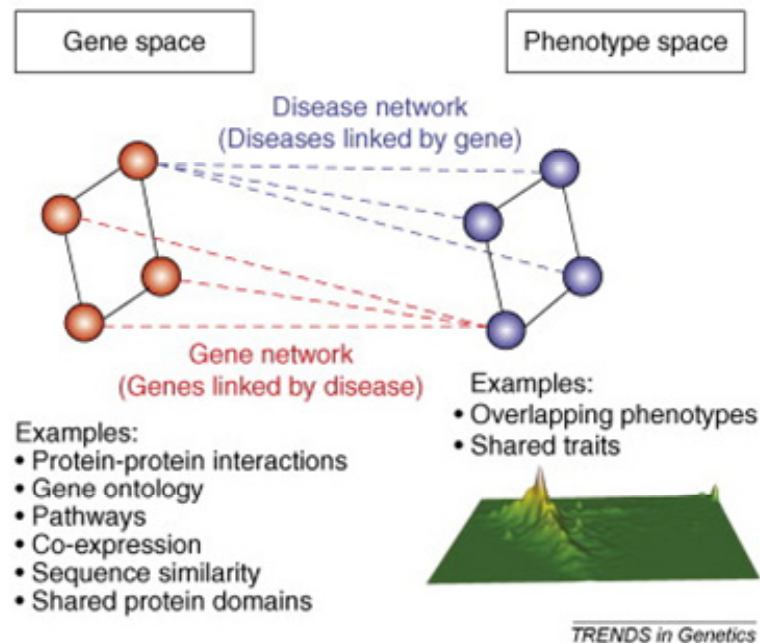
### **V.B.8 Réseau phénotypique et réseau protéique**

De façon intéressante, le réseau de maladies ou d'anomalies observées correspondent à une genèse par une voie de signalisation commune ou par des interactions protéiques communes (Walhout, Reboul et al. 2002; Zhou 2004; Lage, Karlberg et al. 2007; Sam, Liu et al. 2007; Navlakha and Kingsford 2010). Ainsi, différents phénotypes (par exemple des syndromes malformatifs) sont reliés entre eux dans un même réseau phénotypique (ou réseau des maladies) si ils sont causés par un même gène, ou une même famille de gènes, ou des protéines interagissant entre elles, ou des protéines aux fonctions voisines, ou à la structure voisine, ou aux domaines protéiques voisins. C'est ainsi qu'un réseau phénotype-génome peut être dessiné (Butte and Kohane 2006; Oti, Huynen et al. 2008).

---

<sup>48</sup> <http://diseasome.eu/poster.html>

## Relationships between gene space and phenotype space



**Figure 4 :** Interactions entre un réseau de maladies et un réseau de gènes ou de protéines : le réseau phéno- génome. D'après Oti, M., M. A. Huynen, et al. (2008). "Phenome connections." Trends Genet. 24(3):103-106.

### V.C. Phéno- génome et nosologie

En médecine, la méthode analytique a découpé les phénotypes causés par les différentes variations génomiques (CNVs, SNPs, mutations) et leurs interactions avec l'environnement en un ensemble d'entités discrètes, les « maladies », classées selon leurs causes ou leurs principales manifestations. Victor McKusick a ainsi compilé des milliers de maladies génétiques humaines et leurs manifestations en leur donnant un numéro ontologique (numéro de MIM). D'accès libre, OMIM contient actuellement 5000 entrées, dont 2000 sont associées à un gène donné (Amberger, Bocchini et al. 2009).

La phénomique va changer la façon dont nous concevons les maladies génétiques (la nosologie des maladies génétiques), en définissant des « endophénotypes », catégories

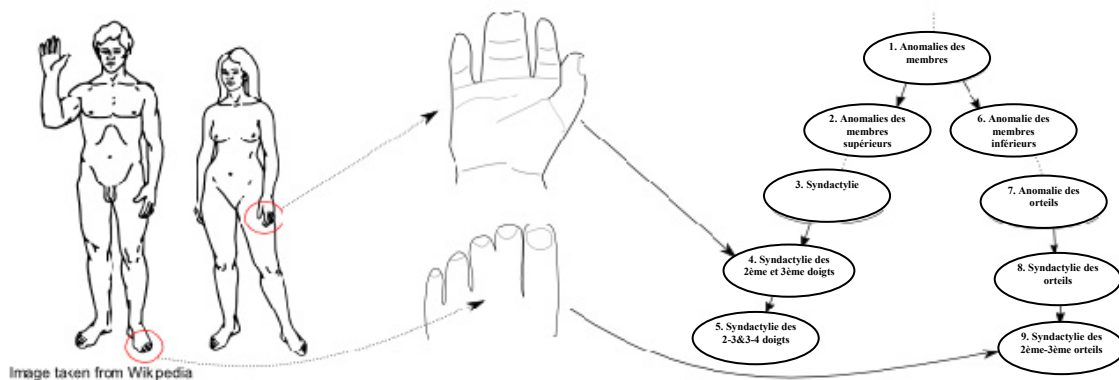


homogènes génétiquement, et en révélant des liens inattendus entre les maladies génétiques (Rzhetsky, 2007 ; Oti M, 2008).

### V.C.1. Une ontologie des phénotypes humains

Lorsque nous avons débuté ce travail en 2007, il n’existait pas d’ontologie des caractères phénotypiques humains ou des anomalies développementales humaines. Cette absence de langage normalisé et contrôlé pour décrire les anomalies développementales est une cause du retard de l’informatisation de l’étude des malformations humaines (Oti, 2008). C’est pourquoi nous avons débuté notre propre travail ontologique.

En 2008, Peter Robinson et al. ont publié une ontologie des phénotypes humains (HPO, Human Phenotype Ontology) développée à partir des synopsis cliniques de OMIM (Robinson, Kohler et al. 2008; Robinson and Mundlos 2010).



**Figure 5 :** Une ontologie des phénotypes humains.

1-Anomalies des membres, 2- Anomalies des membres supérieures, 3-Syndactylie, 4-Syndactylie des 2ème-3ème doigts, 5- Syndactylie des 2-3&3-4 doigts, 6-Anomalies des membres inférieurs, 7-Anomalies des orteils, 8-Syndactylie des orteils, 9-Syndactylie des 2ème-3ème orteils.

Elle contient de nombreuses entrées malformatives mais reste très incomplète sur ce thème (Robinson, Kohler et al. 2008; Robinson and Mundlos 2010). En particulier, elle reste peu détaillée (sa « granularité » reste faible) pour les anomalies malformatives. Les annotations phénotypiques de HPO issues de OMIM (OMIM-HPO) peuvent être enrichies par les entrées de la base de données australienne POSSUM (Oti, Huynen et al. 2009).

Cette faiblesse de l'ontologie HPO entraîne une « sous- annotation » des descriptions cliniques et d'importantes pertes d'informations (Biesecker 2005). Ces améliorations seront nécessaires pour intégrer les données à venir de la génomique personnelle et bien corrélérer variation phénotypique et variation génotypique. Pour cela, notre groupe a contacté le groupe de Peter Robinson, au Centre de bioinformatique de l'Hôpital de la Charité à Berlin. Nous devrions participer à l'enrichissement de l'ontologie en entrées malformatives.

### **V.C.2. Aide au diagnostic**

Un de nos objectifs est de fournir une aide au diagnostic des syndromes malformatifs. Un requérant pourra entrer les différents éléments phénotypiques observés dans un cas autopsique ou clinique. Après des calculs de distance phénotypique, la base de données MDB pourra placer ce cas parmi les cas déjà entrés dans un dendrogramme semblable à celui généré par MimMiner (voir annexe 7- pages xxiii, xxiv). Dans ce cas, l'application ne compare pas des descriptions de syndromes OMIM mais des patients déjà observés. De même, la base MDB pourra comparer ce phénotype avec des cas décrits dans la littérature ou des syndromes OMIM.

Les données statistiques de MDB faciliteront cette aide au diagnostic clinique en hiérarchisant les réponses données, ce que ne permettent pas de faire les requêtes OMIM ou Orphanet à but diagnostique.

### **V.C.3. Description de nouvelles combinaisons malformatives**

Nous ignorons si cette étude permettra de trouver de nouvelles associations syndromiques qui pourront faire l'objet d'une description comme entité clinique. Elle devrait cependant mettre en évidence de nouvelles agrégations phénotypiques, et formuler de nouvelles hypothèses physiopathologiques. En particulier, cette base de données pourra proposer une liste de gènes ou de voies de signalisation candidats, en fonction des malformations ou des combinaisons observées.

### **V.C.4. Visualisation des données**

Les fœtus décrits dans notre base de données seront regroupés automatiquement par proximité phénotypique (agrégation phénotypique) pour réaliser des dendrogrammes semblables à ceux décrits dans MimMiner<sup>49</sup> (voir annexe 7- pages xxiii, xxiv). Pour cela, notre groupe a établi une collaboration avec le « Centre for Molecular and Biomolecular Informatics » de la Radboud University de Nijmegen, aux Pays-bas.

De même, grâce au logiciel Cytoscape<sup>50</sup>, ces cas saisis nous permettront de dessiner un réseau des malformations, comme ce qui a été décrit à propos du « diseasesome » (voir annexe 6, page xxii). Pour cela, nous avons établi une collaboration avec le Dr Lazlo Barabasi<sup>51</sup> de l'Université North Eastern de Boston. Nous verrons ensuite s'il est possible de corrélérer ces réseaux phénotypiques avec des réseaux moléculaires sous-jacents.

---

<sup>49</sup> <http://www.cmbi.ru.nl/MimMiner/>

<sup>50</sup> <http://www.cytoscape.org/>

<sup>51</sup> <http://www.barabasilab.com/>

### **V.C.5. Corrélations génotype phénotype**

Les Drs Zoha Kibar et Jacques Michaud du Centre de recherche du CHU Sainte-Justine ont débuté un travail portant sur les anomalies génomiques des fœtus malformés. Grâce à la base MDB, nous avons pu sélectionner 153 cas de fœtus malformés pour lesquels nous disposons de prélèvements congelés et conservés à -80 °C. Ces fœtus feront l'objet d'une étude génomique par micropuce d'hybridation génomique comparative (aCGH).

Pour cette étude, nous avons choisi les cas dont l'étiologie demeurerait indéterminée en excluant les anomalies chromosomiques et les étiologies génétiques clairement identifiés.

Les anomalies chromosomiques observées seront répertoriées dans la base de données MDB. Elles permettront d'établir rapidement des corrélations génotype- phénotype.

De même, les résultats des études de séquençage haut débit de l'exome ou du génome à venir pourront être intégrés à la base de données.

## **Conclusion**

En conclusion, notre travail de maîtrise comprend la mise en œuvre de trois thésaurus relatifs aux malformations humaines et une base de données phénotypiques spécialisée dans la description des malformations fœtales.

Nous voulons maintenant élargir le recrutement de phénotypes à d'autres centres de fœtopathologie.

Cette base de données servira de base à des applications de phénomique comme :

- la définition d'agrégats phénotypiques (phenotypic clusters) groupant des patients de phénotypes voisins.
- la mesure de distance phénotypique
- la création de représentation en réseau des malformations humaines
- la corrélation de ces réseaux avec des réseaux moléculaires

## **Bibliographie**

- (2002). [No authors listed], "Risks for birth defects or low birth weight: air pollution...medications...and caffeine." Child Health Alert **20**:2. PMID: 12132425 [PubMed - indexed for MEDLINE]
- (2002). [No authors listed]"Urban air pollution linked to birth defects." J Environ Health **65**(2): 47-48. [http://findarticles.com/p/articles/mi\\_hb6679/is\\_2\\_65/ai\\_n28941782/](http://findarticles.com/p/articles/mi_hb6679/is_2_65/ai_n28941782/)
- Amberger, J., C. A. Bocchini, et al. (2009). "McKusick's Online Mendelian Inheritance in Man (OMIM)." Nucleic Acids Res **37**(Database issue): D793-796.
- Ashburner, M., C. A. Ball, et al. (2000). "Gene ontology: tool for the unification of biology. The Gene Ontology Consortium." Nat Genet **25**(1): 25-29.
- Ayme, S. (2000). "Bridging the gap between molecular genetics and metabolic medicine: access to genetic information." Eur J Pediatr **159 Suppl 3**: S183-185.
- Ayme, S. and J. Schmidtke (2007). "Networking for rare diseases: a necessity for Europe." Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz **50**(12): 1477-1483.
- Badano, J. L., N. Mitsuma, et al. (2006). "The ciliopathies: an emerging class of human genetic disorders." Annu Rev Genomics Hum Genet **7**: 125-148.
- Bankier, A. and C. G. Keith (1989). "POSSUM: the microcomputer laser-videodisk syndrome information system." Ophthalmic Paediatr Genet **10**(1): 51-52.
- Barabasi, A. L. (2007). "Network medicine--from obesity to the "diseasome"." N Engl J Med **357**(4): 404-407.
- Baraitser, M. and R. Winter (1991). "Update on the London Dysmorphology Database." Cleft Palate Craniofac J **28**(3): 318.
- Barsnes, H., R. G. Cote, et al. (2010). "OLS dialog: an open-source front end to the ontology lookup service." BMC Bioinformatics **11**: 34.
- Biesecker, L. G. (2005). "Mapping phenotypes to language: a proposal to organize and standardize the clinical descriptions of malformations." Clin Genet **68**(4): 320-326.
- Bilder, R. M., F. W. Sabb, et al. (2009). "Phenomics: the systematic study of phenotypes on a genome-wide scale." Neuroscience **164**(1): 30-42.
- Bilder, R. M., F. W. Sabb, et al. (2009). "Cognitive ontologies for neuropsychiatric phenomics research." Cogn Neuropsychiatry **14**(4-5): 419-450.
- Bodenreider, O. and A. Burgun (2010). "A framework for comparing phenotype annotations of orthologous genes." Stud Health Technol Inform **160**(Pt 2): 1309-1313.
- Bodenreider, O., J. A. Mitchell, et al. (2005). "Biomedical ontologies." Pac Symp Biocomput: 76-78.



- Bogue, M. A. and S. C. Grubb (2004). "The Mouse Phenome Project." Genetica **122**(1): 71-74.
- Bogue, M. A., S. C. Grubb, et al. (2007). "Mouse Phenome Database (MPD)." Nucleic Acids Res **35**(Database issue): D643-649.
- Boyle, E., H. Johnson, et al. (2004). "Congenital anomalies and proximity to landfill sites." Ir Med J **97**(1): 16-18.
- Brandl, J. and T. Grimm (1987). "A chromosome supplement to the London Dysmorphology Database." J Med Genet **24**(8): 497-498.
- Browne, M. L., S. A. Rasmussen, et al. (2009). "Maternal thyroid disease, thyroid medication use, and selected birth defects in the National Birth Defects Prevention Study." Birth Defects Res A Clin Mol Teratol **85**(7): 621-628.
- Brunner, H. G. and M. A. van Driel (2004). "From syndrome families to functional genomics." Nat Rev Genet **5**(7): 545-551.
- Burgun, A., F. Mougin, et al. (2009). "Two approaches to integrating phenotype and clinical information." AMIA Annu Symp Proc **2009**: 75-79.
- Butte, A. J. and I. S. Kohane (2006). "Creation and implications of a phenome-genome network." Nat Biotechnol **24**(1): 55-62.
- Chudley, A. E., A. R. Kilgour, et al. (2007). "Challenges of diagnosis in fetal alcohol syndrome and fetal alcohol spectrum disorder in the adult." Am J Med Genet C Semin Med Genet **145C**(3): 261-272.
- Cote, R., F. Reisinger, et al. (2010). "The Ontology Lookup Service: bigger and better." Nucleic Acids Res **38** **Suppl**: W155-160.
- Cote, R. G., P. Jones, et al. (2006). "The Ontology Lookup Service, a lightweight cross-platform tool for controlled vocabulary queries." BMC Bioinformatics **7**: 97.
- Cote, R. G., P. Jones, et al. (2008). "The Ontology Lookup Service: more data and better tools for controlled vocabulary queries." Nucleic Acids Res **36**(Web Server issue): W372-376.
- de la Cruz, N., S. Bromberg, et al. (2005). "The Rat Genome Database (RGD): developments towards a phenome database." Nucleic Acids Res **33**(Database issue): D485-491.
- Dodds, L. and W. D. King (2001). "Relation between trihalomethane compounds and birth defects." Occup Environ Med **58**(7): 443-446.

- Dolk, H. (2004). "Epidemiologic approaches to identifying environmental causes of birth defects." Am J Med Genet C Semin Med Genet **125C**(1): 4-11.
- Elliott, P., S. Richardson, et al. (2009). "Geographic density of landfill sites and risk of congenital anomalies in England." Occup Environ Med **66**(2): 81-89.
- Ferencz, C., C. Loffredo, et al. (2000). "Epidemiologic methods in the search for causes of human birth defects." Teratology **61**(4): 243-245.
- Fernandez-Ricaud, L., J. Warringer, et al. (2007). "PROPHECY--a yeast phenome database, update 2006." Nucleic Acids Res **35**(Database issue): D463-467.
- Fitch, N. (1977). "Syndromology." J Pediatr **90**(5): 852-853.
- Freimer, N. and C. Sabatti (2003). "The human phenome project." Nat Genet **34**(1): 15-21.
- Fryns, J. P. and T. J. de Ravel (2002). "London Dysmorphology Database, London Neurogenetics Database and Dysmorphology Photo Library on CD-ROM [Version 3] 2001R. M. Winter, M. Baraitser, Oxford University Press, ISBN 019851-780, pound sterling 1595." Hum Genet **111**(1): 113.
- Gkoutos, G. V., C. Mungall, et al. (2009). "Entity/quality-based logical definitions for the human skeletal phenome using PATO." Conf Proc IEEE Eng Med Biol Soc **2009**: 7069-7072.
- Goh, K. I., M. E. Cusick, et al. (2007). "The human disease network." Proc Natl Acad Sci U S A **104**(21): 8685-8690.
- Gold, W. S. (1990). "Is there a connection between pollution and birth defects?" Tex Med **86**(1): 10.
- Groth, P., I. Kalev, et al. (2010). "Phenoclustering: online mining of cross-species phenotypes." Bioinformatics **26**(15): 1924-1925.
- Groth, P., N. Pavlova, et al. (2007). "PhenomicDB: a new cross-species genotype/phenotype resource." Nucleic Acids Res **35**(Database issue): D696-699.
- Grubb, S. C., T. P. Maddatu, et al. (2009). "Mouse phenome database." Nucleic Acids Res **37**(Database issue): D720-730.
- Guest, S. S., C. D. Evans, et al. (1999). "The Online London Dysmorphology Database." Genet Med **1**(5): 207-212.
- Hall, Judith G. (1992). "Developmental Defects in Stillborn and Newborn Infants." (Figure 5-12):132. Developmental Pathology of the Embryo and Fetus. James E. Dimmick et Dagmar K. Kalousek. Philadelphia, J.B. Lippincott.

- Hansen, C. A., A. G. Barnett, et al. (2009). "Ambient air pollution and birth defects in brisbane, australia." PLoS One **4**(4): e5408.
- Hillman, S. C., S. Pretlove, et al. (2010). "Additional information from array Comparative Genomic Hybridisation (array CGH) technology over conventional karyotyping in prenatal diagnosis - a systematic review and meta-analysis." Ultrasound Obstet Gynecol.
- Hsu, C. H., S. P. Lin, et al. (1994). "Computerized Dysmorphology Database Mackay (CDDM): preliminary report." Zhonghua Min Guo Xiao Er Ke Yi Xue Hui Za Zhi **35**(1): 19-26.
- Jaqaman, K., J. F. Dorn, et al. (2007). "Phenotypic clustering of yeast mutants based on kinetochore microtubule dynamics." Bioinformatics **23**(13): 1666-1673.
- Jiang, X., B. Liu, et al. (2008). "Modularity in the genetic disease-phenotype network." FEBS Lett **582**(17): 2549-2554.
- Johnson, L. A., L. L. Barton, et al. (1981). "Clustering of phenotypic variables and construction of an identification scheme for human isolates of Haemophilus." J Infect **3**(1): 59-70.
- Kahraman, A., A. Avramov, et al. (2005). "PhenomicDB: a multi-species genotype/phenotype database for comparative phenomics." Bioinformatics **21**(3): 418-420.
- Khoury, M. J., J. E. Becerra, et al. (1989). "Maternal thyroid disease and risk of birth defects in offspring: a population-based case-control study." Paediatr Perinat Epidemiol **3**(4): 402-420.
- Kloppenborg, S., U. K. Brandt, et al. (2005). "Risk of congenital anomalies in the vicinity of waste landfills in Denmark; an epidemiological study using GIS." Cent Eur J Public Health **13**(3): 137-143.
- Klotz, J. B. and L. A. Pyrch (1999). "Neural tube defects and drinking water disinfection by-products." Epidemiology **10**(4): 383-390.
- Knowlton, M. N., T. Li, et al. (2008). "A PATO-compliant zebrafish screening database (MODB): management of morpholino knockdown screen information." BMC Bioinformatics **9**: 7.
- Kriek, M., S. J. White, et al. (2006). "Copy number variation in regions flanked (or unflanked) by duplicons among patients with developmental delay and/or congenital malformations; detection of reciprocal and partial Williams-Beuren duplications." Eur J Hum Genet **14**(2): 180-189.
- Krous, H. F. (1981). "Fetal alcohol syndrome: a dilemma of maternal alcoholism." Pathol Annu **16 Pt 1**: 295-311.

- Kumar, S. (2002). "Pesticide ban imposed after reports of congenital abnormalities." BMJ **325**(7360): 356.
- Kuromori, T., T. Wada, et al. (2006). "A trial of phenome analysis using 4000 Ds-insertional mutants in gene-coding regions of Arabidopsis." Plant J **47**(4): 640-651.
- Lage, K., E. O. Karlberg, et al. (2007). "A human phenome-interactome network of protein complexes implicated in genetic disorders." Nat Biotechnol **25**(3): 309-316.
- Lin, S., S. A. Hwang, et al. (1998). "Does paternal occupational lead exposure increase the risks of low birth weight or prematurity?" Am J Epidemiol **148**(2): 173-181.
- Loeken, M. R. (2006). "Advances in understanding the molecular causes of diabetes-induced birth defects." J Soc Gynecol Investig **13**(1): 2-10.
- Luke, B. (1977). "Maternal alcoholism and fetal alcohol syndrome." Am J Nurs **77**(12): 1924-1926.
- Lussier, Y., T. Borlawsky, et al. (2006). "PhenoGO: assigning phenotypic context to gene ontology annotations with natural language processing." Pac Symp Biocomput: 64-75.
- Lussier YA, Liu Y. (2007). "Computational Approaches to Phenotyping." Proc Am Thorac Soc **4**(1):18-25.
- Maier, H., C. Lengger, et al. (2008). "MausDB: an open source application for phenotype data and mouse colony management in large-scale mouse phenotyping projects." BMC Bioinformatics **9**: 169.
- Marquet, G., J. Mosser, et al. (2007). "A method exploiting syntactic patterns and the UMLS semantics for aligning biomedical ontologies: the case of OBO disease ontologies." Int J Med Inform **76 Suppl 3**: S353-361.
- Martinez-Frias, M. L., J. L. Frias, et al. (1998). "Errors of morphogenesis and developmental field theory." Am J Med Genet **76**(4): 291-296.
- Masuya, H., S. Yoshikawa, et al. (2007). "Phenosite: a web database integrating the mouse phenotyping platform and the experimental procedures in mice." J Bioinform Comput Biol **5**(6): 1173-1191.
- McKusick, V. A. (1969). "On lumpers and splitters, or the nosology of genetic disease." Perspect Biol Med **12**(2): 298-312.
- Midic, U., C. J. Oldfield, et al. (2009). "Unfoldomics of human genetic diseases: illustrative examples of ordered and intrinsically disordered members of the human diseasome." Protein Pept Lett **16**(12): 1533-1547.

- Monteleone-Neto, R., D. Brunoni, et al. (1985). "Birth defects and environmental pollution: the Cubatao example." Prog Clin Biol Res **163B**: 65-68.
- Muilu, J., L. Peltonen, et al. (2007). "The federated database--a basis for biobank-based post-genome studies, integrating phenome and genome data from 600,000 twin pairs in Europe." Eur J Hum Genet **15**(7): 718-723.
- Navlakha, S. and C. Kingsford (2010). "The power of protein interaction networks for associating genes with diseases." Bioinformatics **26**(8): 1057-1063.
- Nelson, K. and L. B. Holmes (1989). "Malformations due to presumed spontaneous mutations in newborn infants." N Engl J Med **320**(1): 19-23.
- Neumann, P. E., W. N. Frankel, et al. (1994). "Multifactorial inheritance of neural tube defects: localization of the major gene and recognition of modifiers in ct mutant mice." Nat Genet **6**(4): 357-362.
- Oti, M. and H. G. Brunner (2007). "The modular nature of genetic diseases." Clin Genet **71**(1): 1-11.
- Oti, M., M. A. Huynen, et al. (2008). "Phenome connections." Trends Genet. **24**(3):103-106.
- Oti, M., M. A. Huynen, et al. (2009). "The biological coherence of human phenome databases." Am J Hum Genet **85**(6): 801-808.
- Oti, M., B. Snel, et al. (2006). "Predicting disease genes using protein-protein interactions." J Med Genet **43**(8): 691-698.
- Paigen, K. and J. T. Eppig (2000). "A mouse phenome project." Mamm Genome **11**(9): 715-717.
- Palmer, S. R., F. D. Dunstan, et al. (2005). "Risk of congenital anomalies after the opening of landfill sites." Environ Health Perspect **113**(10): 1362-1365.
- Pelz, J., V. Arendt, et al. (1996). "Computer assisted diagnosis of malformation syndromes: an evaluation of three databases (LDDb, POSSUM, and SYNDROC)." Am J Med Genet **63**(1): 257-267.
- Perez-Iratxeta, C., P. Bork, et al. (2002). "Association of genes to genetically inherited diseases using data mining." Nat Genet **31**(3): 316-319.
- Potash, J. B., J. Toolan, et al. (2007). "The bipolar disorder phenome database: a resource for genetic studies." Am J Psychiatry **164**(8): 1229-1237.
- Quinlan, R. J., J. L. Tobin, et al. (2008). "Modeling ciliopathies: Primary cilia in development and disease." Curr Top Dev Biol **84**: 249-310.

- Ritz, B., F. Yu, et al. (2002). "Ambient air pollution and risk of birth defects in Southern California." Am J Epidemiol **155**(1): 17-25.
- Robinson, P. N., S. Kohler, et al. (2008). "The Human Phenotype Ontology: a tool for annotating and analyzing human hereditary disease." Am J Hum Genet **83**(5): 610-615.
- Robinson, P. N. and S. Mundlos (2010). "The human phenotype ontology." Clin Genet **77**(6): 525-534.
- Rzhetsky, A., D. Wajngurt, et al. (2007). "Probing genetic overlap among complex human phenotypes." Proc Natl Acad Sci U S A **104**(28): 11694-11699.
- Sabb, F. W., A. C. Burggren, et al. (2009). "Challenges in phenotype definition in the whole-genome era: multivariate models of memory and intelligence." Neuroscience **164**(1): 88-107.
- Sam, L., Y. Liu, et al. (2007). "Discovery of protein interaction networks shared by diseases." Pac Symp Biocomput: 76-87.
- Sam, L. T., E. A. Mendonca, et al. (2009). "PhenoGO: an integrated resource for the multiscale mining of clinical and biological data." BMC Bioinformatics **10 Suppl 2**: S8.
- Sardana, D., S. Vasa, et al. (2010). "PhenoHM: human-mouse comparative phenome-genome server." Nucleic Acids Res **38 Suppl**: W165-174.
- Schorderet, D. F. (1991). "Using OMIM (On-line Mendelian Inheritance in Man) as an expert system in medical genetics." Am J Med Genet **39**(3): 278-284.
- Shaffer, L. G., B. A. Bejjani, et al. (2007). "The identification of microdeletion syndromes and other chromosome abnormalities: cytogenetic methods of the past, new technologies for the future." Am J Med Genet C Semin Med Genet **145C**(4): 335-345.
- Shaffer, L. G., A. Theisen, et al. (2007). "The discovery of microdeletion syndromes in the post-genomic era: review of the methodology and characterization of a new 1q41q42 microdeletion syndrome." Genet Med **9**(9): 607-616.
- Shaw, G. M., C. R. Wasserman, et al. (1999). "Maternal pesticide exposure from multiple sources and selected congenital anomalies." Epidemiology **10**(1): 60-66.
- Shinawi, M. and S. W. Cheung (2008). "The array CGH and its clinical applications." Drug Discov Today **13**(17-18): 760-770.
- Slavotinek, A. M. (2008). "Novel microdeletion syndromes detected by chromosome microarrays." Hum Genet **124**(1): 1-17.

- Smith, B., M. Ashburner, et al. (2007). "The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration." Nat Biotechnol **25**(11): 1251-1255.
- Smith, C. L. and J. T. Eppig (2009). "The mammalian phenotype ontology: enabling robust annotation and comparative analysis." Wiley Interdiscip Rev Syst Biol Med **1**(3): 390-399.
- Smith, C. L., C. A. Goldsmith, et al. (2005). "The Mammalian Phenotype Ontology as a tool for annotating, analyzing and comparing phenotypic information." Genome Biol **6**(1): R7.
- Spranger, J., K. Benirschke, et al. (1982). "Errors of morphogenesis: concepts and terms. Recommendations of an international working group." J Pediatr **100**(1): 160-165.
- Stankiewicz, P. and A. L. Beaudet (2007). "Use of array CGH in the evaluation of dysmorphology, malformations, developmental delay, and idiopathic mental retardation." Curr Opin Genet Dev **17**(3): 182-192.
- Tobin, J. L. and P. L. Beales (2009). "The nonmotile ciliopathies." Genet Med **11**(6): 386-402.
- Trasler, J. M. and T. Doerksen (1999). "Teratogen update: paternal exposures-reproductive risks." Teratology **60**(3): 161-172.
- van Driel, M. A., J. Bruggeman, et al. (2006). "A text-mining analysis of the human phenome." Eur J Hum Genet **14**(5): 535-542.
- Veltman, J. A. and H. G. Brunner (2010). "Understanding variable expressivity in microdeletion syndromes." Nat Genet **42**(3): 192-193.
- Vrijheid, M., H. Dolk, et al. (2002). "Chromosomal congenital anomalies and residence near hazardous waste landfill sites." Lancet **359**(9303): 320-322.
- Walhout, A. J., J. Reboul, et al. (2002). "Integrating interactome, phenome, and transcriptome mapping data for the *C. elegans* germline." Curr Biol **12**(22): 1952-1958.
- Winter, R. M. and M. Baraitser (1987). "The London Dysmorphology Database." J Med Genet **24**(8): 509-510.
- Wu, X., Q. Liu, et al. (2009). "Align human interactome with phenome to identify causative genes and networks underlying disease families." Bioinformatics **25**(1): 98-104.
- Zhou, H. X. (2004). "Improving the understanding of human genetic diseases through predictions of protein structures and protein-protein interaction sites." Curr Med Chem **11**(5): 539-549.

## **Annexes**



## Annexe 1

### Tableau 2 avec plus de détails

Overview of the Considered Syndrome Databases. Oti, M., M. A. Huynen, et al. (2009).

"The biological coherence of human phenome databases." Am J Hum Genet 85(6): 801-808

		OMIM		
		Orphanet <sup>a</sup>	POSSUM	HPO <sup>b</sup> MimMiner <sup>c</sup>
Num. syndromes		2070	3167	4779 5948
Num. features in ontology		864	1115	8275 1368 <sup>d</sup>
Feature ontology depth: max. (median)		4 (2)	2 (2)	13 (6) 15 (5)
Median num. features per syndrome: original (expanded) <sup>e</sup>		13 (25)	22 (34)	7 (20) 8 (22)
Num. syndromes mapped to disease genes		668	924	2053 2055
Num. disease genes		1038 <sup>f</sup>	986	2019 1937

Feature ontology depth is the level of detail provided to define a phenotype.

<sup>a</sup>Only the feature-annotated Orphanet syndromes were included. There were a total of 7435 syndrome IDs in the full Orphanet database. However, Orphanet uses clinical syndrome definitions, which are broader than the locus-based syndromes in databases such as OMIM. As a result, multiple IDs associated with the same clinical syndrome are not separately annotated with features.

<sup>b</sup>The Human Phenotype Ontology contains phenotype annotation for a subset of OMIM syndromes as well as some OMIM genes.

<sup>c</sup>The MimMiner text-mining conversion of the OMIM database is listed for comparison.

<sup>d</sup>The “Anatomy” (A) and “Pathological Conditions, Signs, and Symptoms” (C23) parts of the MeSH ontology were used in the MimMiner approach.

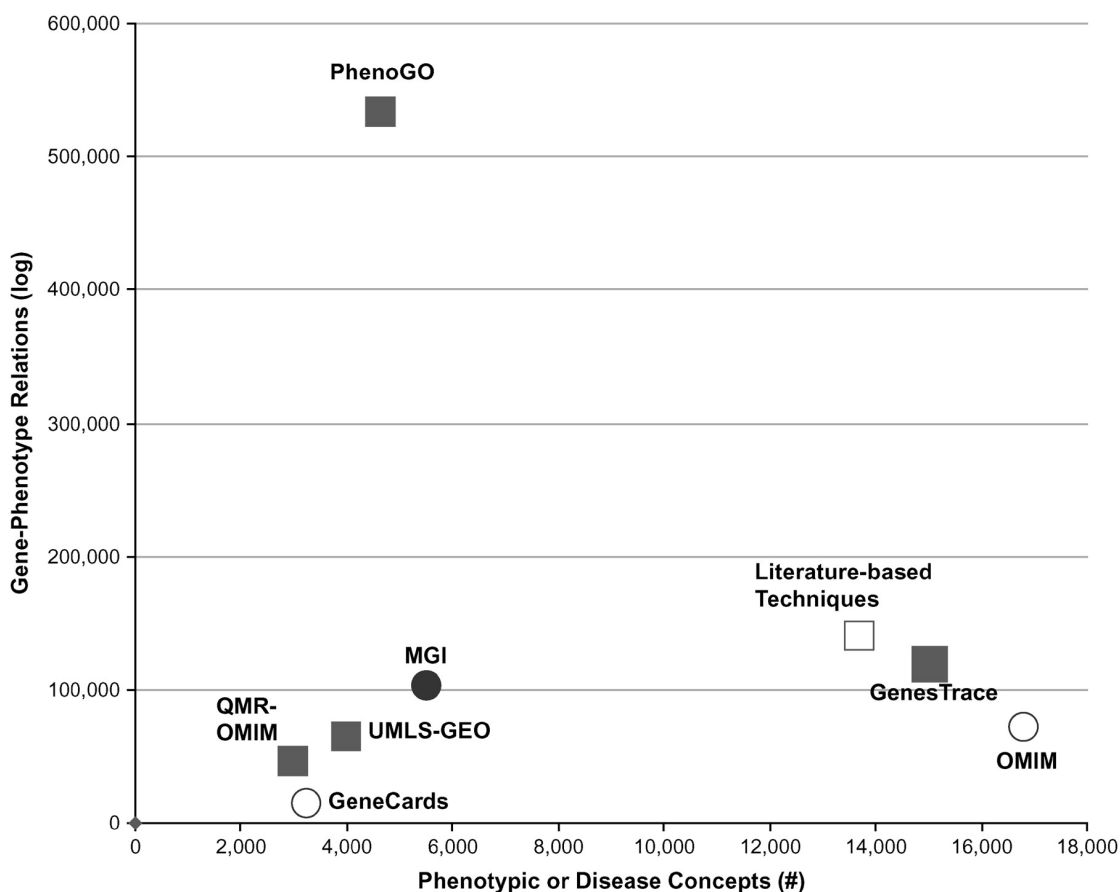
<sup>e</sup>The number in parentheses refers to the median number of features per syndrome after the syndrome feature vectors are expanded to include the feature's ontological ancestors in the feature vector.

<sup>f</sup>The Orphanet database used originally contained 569 disease genes, but this number was expanded to 1935 via syndrome-to-gene mappings from the OMIM database. 1038 of these were associated with feature-annotated syndromes.

## Annexe 2:

Comparaison du nombre de phénotypes et le nombre de relations gène-phénotype dans les bases de données et les réseaux gène- phénotype

Lussier YA, Liu Y. (2007). "Computational Approaches to Phenotyping." *Proc Am Thorac Soc* 4(1):18-25.



Comparison of the number of distinct phenotypes and the number of gene–phenotype relationships in gene–phenotype databases and networks, showing that the PhenoGO is the largest network and that both literature text-mining techniques and OMIM provide the broadest annotations of distinct phenotypes with genes. *Solid circle*, ontology-anchored database; *open circles*, database with unstructured phenotypes; *solid squares*, ontology-anchored high-throughput phenotyping (HTP) phenomics; *open square*, HTP phenomics with unstructured phenotypes.

MGI = Mouse Genome Informatics; OMIM = Online Mendelian Inheritance in Man; QMR = Quick Medical Reference; UMLS-GEO = Unified Medical Language System–Gene Expression Omnibus

### Annexe 3:

Trois syndromes (synopsis cliniques d'OMIM) qui partagent des phénotypes similaires.

(Les phénotypes similaires indiqués en vert).

#### BARDET-BIEDL SYNDROME; BBS (MIM ID# 209900 (600151)) Clinical Synopsis

Heading	Subheading	Feature
Head and neck	Eye	Rod-cone dystrophy, onset by end of 2 <sup>nd</sup> decade (major), Retinitis pigmentosa, Strabismus, Cataracts
	Mouth	High arched palate
	Teeth	Dental crowding Hypodontia, Small tooth
Cardiovascular	Heart	Left ventricular hypertrophy, Congenital heart defects
Abdomen	Liver	Hepatic fibrosis
	Gastrointestinal	Hirschsprung disease (<10%)
Genitourinary	Internal genitalia-male	Hypogonadism
	Kidneys	Renal anomalies (major) Nephrogenic diabetes
Skeletal	Hands	Polydactyly (major), Brachydactyly
	Feet	Polydactyly (major)
Neurologic	Central nervous system	Speech disorder, Speech delay, Learning disabilities (major), Poor coordination, Ataxia, Developmental delay

**BIEMOND SYNDROME II (MIM ID# 210350)**  
Clinical Synopsis

Heading	Subheading	Feature
Head and neck	Eye	Iris coloboma
Skeletal	Hands	Preaxial polydactyly
Neurologic	Central nervous system	Mental retardation
Endocrine features		Hypogonadism

**MCKUSICK-KAUFMAN SYNDROME; MKKS (MIM ID #236700)**  
**HYDROMETROCOLPOS, POSTAXIAL POLYDACTYLY,**  
**AND CONGENITAL HEART MALFORMATION; HMCS**  
Clinical Synopsis

Heading	Subheading	Feature
Cardiovascular	Heart	Congenital heart defects
Respiratory	Lung	Hypoplastic lungs
Abdomen	Gastrointestinal	Hirschsprung disease (%12), Imperforate anus (8%), Rectovaginal fistula
Genitourinary	Internal genitalia, male	Cryptorchidism
	Internal genitalia, female	Hydrometrocolpos
	Kidneys	Polycystic kidney, Hydronephrosis
	Ureters	Hydroureter
	Bladder	Vesicovaginal fistula
Skeletal	Pelvis	Congenital dislocation of the hip
	Hand	Mesoaxial polydactyly, Postaxial polydactyly, Syndactyly
Muscle, soft tissue		Leg edema

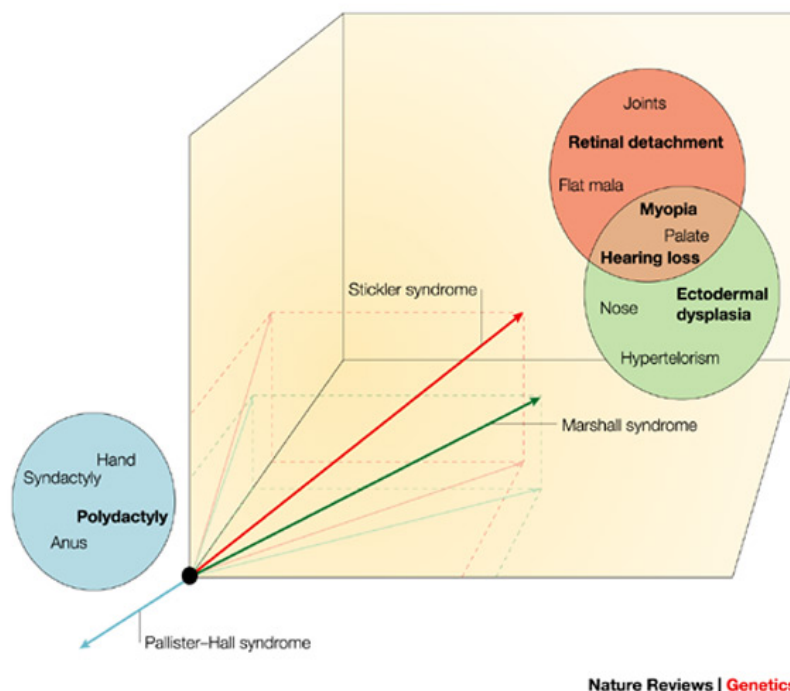
The study by Van Driel et al. [MimMiner] is an attempt to identify a phenotype network by linking multiple disparate phenotypes to single genes. In this study, the shared predisposing genes were inferred from the co-occurrence of diseases in patients, and no genetic studies were performed. 'MimMiner' enables the user to retrieve the similarity ranking for a specific OMIM phenotype.

## Annexe 4:

### Agrégation phénotypique et famille de maladies

(phenotype clustering). Brunner, H. G. and M. A. van Driel (2004). "From syndrome families to functional genomics." *Nat Rev Genet* **5**(7): 545-551.

Each arrow represents a **KEYWORD VECTOR**. The components in a keyword vector correspond to terms in the document (for example, 'myopia', 'joints', and so on). Vectors that point in the same direction are more alike. Three documents that describe three syndromes are shown (red: **STICKLER SYNDROME**; green: **MARSHALL SYNDROME**; and blue: **PALLISTER–HALL SYNDROME**). Stickler and Marshall syndromes have similar phenotypes and share common terms ('myopia', 'palate', 'hearing loss') in their keyword vectors. **Pallister–Hall syndrome** is clearly different, and shares few terms with Stickler and Marshall, and therefore, the vector points in a different direction. Some terms are more important than other terms. Terms are assigned different weights by computing the **INVERSE DOCUMENT FREQUENCY**, which gives a weight factor to each term in the collection. Terms in bold contribute the most to vector direction. Flat mala, low cheekbones; hypertelorism, widely-spaced eyes.



Two of these three syndromes (OMIM clinical synopsis) have similar phenotypes and share common terms. Similar phenotypes are in green.

STICKLER SYNDROME, TYPE I; STL1 (MIM ID #108300)  
Clinical Synopsis

Heading	Subheading	Feature
Head and neck	Face	Flat midface
	Eye	Myopia, Retinal detachment, Blindness, Occasional cataracts
	Mouth	Cleft palate, Pierre-Robin sequence
	Ear	Sensorineural hearing loss, Occasional conductive hearing loss
Cardiovascular	Heart	Mitral valve prolapse
Skeletal		Mild spondyloepiphyseal dysplasia

MARSHALL SYNDROME (MIM ID #154780)  
Clinical Synopsis

Heading	Subheading	Feature
Head and neck	Face	Flat midface, Micrognathia, Long philtrum
	Eye	Myopia, Congenital cataracts, Esotropia, Retinal detachment, Glaucoma, Lens dislocation, Hypertelorism
	Mouth	Cleft palate, Pierre-Robin sequence, Thick lips
	Ear	Sensorineural hearing loss, Low set ear
Skeletal	Skull	Calvarial thickening, Absent frontal sinuses, Falx, tentorial, and meningeal calcifications
	Spine	Mild platyspondyly
	Pelvis	Small iliac bones, Coxa valga
	Limbs	Small, irregular distal femoral epiphyses, Small, irregular proximal tibial epiphyses, Outward radial bowing, Outward ulnar bowing
	Hands	Wide tufts of distal phalanges

PALLISTER-HALL SYNDROME; PHS.(MIM ID #146510)  
Clinical Synopsis

Heading	Subheading	Feature
Head and neck	Eye	Microphthalmia
	Mouth	Multiple buccal frenula, Microglossia, Cleft lip and palate
	Ear	Simple auricles, Absent external auditory canals, Microtia
Cardiovascular	Heart	Ventricular septal defect, Proximal aortic coarctation
Respiratory	Larynx	Laryngeal cleft, Bifid or hypoplastic epiglottis
	Lung	Abnormal lung lobation
Chest	Ribs, sternum, clavicles, and scapulae	Fused rib
Abdomen	Gastrointestinal	Imperforate anus
Skeletal	Spine	Hemivertebrae
	Pelvis	Dislocated hips
	Limbs	Distal shortening of limbs, Radial subluxation

## Annexe 5

### Quelques exemples des données statistiques de la base de données MDB

(<http://www.malformations.org/Thesaurus/view/160>)

Associated developmental anomalies.

Here is the list of developmental anomalies associated with macrostomia, DA: 160, based on the cases' informations :

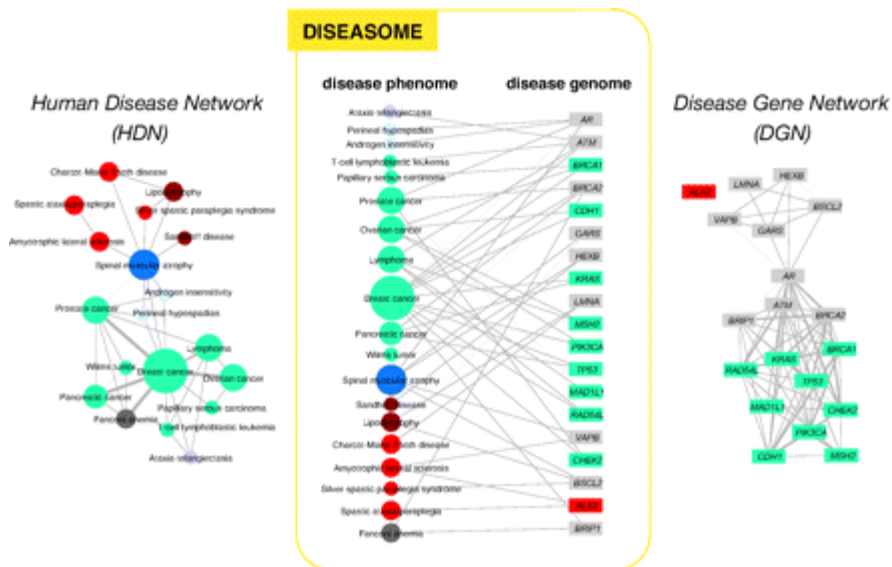
- [Large nose, fetal](#) 9.09% (2/22 anomalies), in cases [MDB395](#), [MDB543](#)
- [Facial dysmorphism](#) 9.09% (2/22 anomalies), in cases [MDB543](#), [MDB315](#)
- [Placenta hypoplasia](#) 9.09% (2/22 anomalies), in cases [MDB543](#), [MDB315](#)
- [Fragile X Syndrome](#) 4.55% (1/22 anomalies), in case [MDB395](#)
- [Gothic palate](#) 4.55% (1/22 anomalies), in case [MDB395](#)
- [Large ear, fetal](#) 4.55% (1/22 anomalies), in case [MDB395](#)
- [Triploidy, Maternal, nonmolar \(triploidy by digyny\)](#) 4.55% (1/22 anomalies), in case [MDB543](#)
- [Arthrogryposis](#) 4.55% (1/22 anomalies), in case [MDB543](#)
- [Low-set ears](#) 4.55% (1/22 anomalies), in case [MDB543](#)
- [Retrognathia](#) 4.55% (1/22 anomalies), in case [MDB543](#)
- [Large anterior fontanelle](#) 4.55% (1/22 anomalies), in case [MDB543](#)
- [Neck edema](#) 4.55% (1/22 anomalies), in case [MDB543](#)
- [Retroplacental hemorrhage](#) 4.55% (1/22 anomalies), in case [MDB543](#)
- [Stenosis of the aqueduct of Sylvius](#) 4.55% (1/22 anomalies), in case [MDB315](#)
- [Microcephaly](#) 4.55% (1/22 anomalies), in case [MDB315](#)
- [Clindodactyly of 5th finger, bilateral](#) 4.55% (1/22 anomalies), in case [MDB315](#)
- [Cerebellar hypoplasia](#) 4.55% (1/22 anomalies), in case [MDB315](#)
- [Hypertelorism](#) 4.55% (1/22 anomalies), in case [MDB315](#)
- [Micrognathia](#) 4.55% (1/22 anomalies), in case [MDB315](#)



## Annexe 6

### Diseasome

Goh, Cusick et al. 2007. “The human disease network”. Proc Natl Acad Sci U S A. 22;104(21):8685-90.



Construction of the diseasome bipartite network. (*Center*) A small subset of OMIM-based disorder–disease gene associations where circles and rectangles correspond to disorders and disease genes, respectively. A link is placed between a disorder and a disease gene if mutations in that gene lead to the specific disorder. The size of a circle is proportional to the number of genes participating in the corresponding disorder, and the color corresponds to the disorder class to which the disease belongs. (*Left*) The HDN projection of the diseasome bipartite graph, in which two disorders are connected if there is a gene that is implicated in both. The width of a link is proportional to the number of genes that are implicated in both diseases. For example, three genes are implicated in both breast cancer and prostate cancer, resulting in a link of weight three between them. (*Right*) The DGN projection where two genes are connected if they are involved in the same disorder. The width of a link is proportional to the number of diseases with which the two genes are commonly associated.

## Annexe 7:

### Quelques cas (fœtus) de notre base de données MDB

Ces fœtus semblables aux syndromes décrits dans OMIM (synopsis cliniques) et feraient l'objet d'approche MimMiner.

(Les phénotypes similaires sont indiqués en vert)

#### Case (foetus) HSJ14

- *Holoprosencephaly*, alobar
- *Microcephaly*
- *Corpus callosum agenesis*
- Thalamic nuclei centrum medianum fusion
- *Absence of the pyramid tracts, bilateral*
- Hypogenesis of the Cerebellar Vermis
- Hypotelorism
- Obstetrician's hand
- Polymicrogyria, focal
- *Cleft palate with cleft lip, bilateral, complete*

#### Case (foetus) HSJ15

- *Microcephaly*
- Microtia
- *Cleft palate with cleft lip*, unilateral, complete
- *Overlapping fingers*
- Renal agenesis, unilateral
- Intrauterine growth retardation, symmetric
- Placenta hypoplasia
- Cerebellar hypoplasia
- Cerebral ventriculomegaly

#### Case (foetus) HSJ166

- Facial dysmorphism
- *Hypertelorism*
- *Cleft palate with cleft lip (CLP)*, median
- Exencephaly
- Gastroschisis
- Fingers amputation
- Clindodactyly of 5th finger, unilateral
- Semilobar *holoprosencephaly*
- Optic nerve, absence of, bilateral
- Olfactory bulbs, absence of
- *Corpus callosum agenesis*
- *Microcephaly*

**Case (foetus) HSJ175**

- *Holoanencephaly*
- Rachischisis, cervical
- Anencephaly facies
- *Microcephaly*
- Exophthalmos, bilateral
- Large nose, fetal
- Short neck
- *Overlapping fingers*
- Pulmonary hypoplasia
- Adrenal hypoplasia, fetal
- Macroglossia
- Ureteral duplication, complete, unilateral
- Badly-hemmed ears
- Low-set ears

These foetuses sharing the key features resembling “family syndromes-syndromes within syndromes” and share common terms suitable for MimMiner approach.